THE ART OF SIMULATION

ELECTRICAL ENGINEERING SERIES

General Editor

ARTHUR PORTER,

PH.D., M.I.E.E.

Professor of Industrial Engineering, University of Toronto;
Formerly Dean of Engineering,
University of Saskatchewan, Canada;
Professor of Light Electrical Engineering,
Imperial College of Science and Technology, London

ANALYTICAL TECHNIQUES
FOR NON-LINEAR CONTROL SYSTEMS

J. C. WEST,

D.SC., A.M.I.E.E., A.M.A.I.E.E.

AN INTRODUCTION TO
LINEAR NETWORK ANALYSIS

P. S. FARAGO,

DR.PHIL., F.INST.P., F.R.S.E.

HIGH FREQUENCY APPLICATION OF FERRITES

J. ROBERTS,

B.SC. (ENG.), A.C.G.I., A.M.I.E.E.

EXAMPLES IN
ADVANCED ELECTRICAL ENGINEERING

J. F. W. BELL,

B.SC., PH.D., A.M.I.E.E.

THE ART OF SIMULATION

K. D. TOCHER
B.Sc., Ph.D., D.Sc.

THE ENGLISH UNIVERSITIES PRESS LTD
102 NEWGATE STREET
LONDON · EC1

First Printed 1963

To
Charlotte

PRINTED AND BOUND IN ENGLAND
FOR THE ENGLISH UNIVERSITIES PRESS LTD
BY HAZELL WATSON AND VINEY LTD. AYLESBURY, BUCKS

GENERAL EDITOR'S FOREWORD

by

ARTHUR PORTER, PH.D., M.I.E.E.

Professor of Industrial Engineering, University of Toronto;
Formerly Dean of Engineering, University of Saskatchewan, Canada;
Professor of Light Electrical Engineering,
Imperial College of Science and Technology, London

This volume is one of a series of texts in electrical engineering, and related subjects, the main object of which is to present advances in these subjects on the one hand, and to re-evaluate basic principles, in the light of recent developments, on the other. The books will be of special interest to research workers, postgraduates, and advanced undergraduates in electrical science and applied physics.

The publication of such a series is regarded as timely because of the increasing realisation that the average three or four years' engineering degree course is inadequate for the basic training of research engineers. To correct this situation there has been a notable increase, during the past few years, of postgraduate courses in all branches of engineering in the Universities and Colleges of Advanced Technology of the United Kingdom, the British Commonwealth, and the United States. Strong official support for these courses is being given by the Department of Scientific and Industrial Research, London, and their importance cannot be over-emphasised. But such specialised courses require suitable texts and many of them are seriously handicapped in this respect. This is not surprising in view of the proliferation of science and technology and the recognition that new languages—one of which is the language of high-speed digital computers—have been born within the last decade. It is essential, moreover, that the authors of texts at advanced levels should be actively engaged in research work because only the active participant working at the frontiers of his subject is in a position to assess the importance of new concepts and of the mathematical and experimental techniques involved. I believe that this new series of electrical engineering texts will fulfil a real need and that it will be welcomed as a medium for re-evaluating fundamental engineering principles in the light of recent advances with special emphasis on the needs of postgraduate students.

The present volume deals with the study of industrial operations and processes using large-scale data-processing and computer systems as simulators. It is perhaps the first book to be published on the subject. During recent years considerable advances have been made in the development of models of technological and sociological systems, especially those embodying information feedback, but a major difficulty has been the handling of the extensive computational work. Dr. Tocher has made notable contributions to both statistical mathematics, and to the art of simulation, especially in areas involving a deep understanding of probability theory. This book will be particularly welcomed by operational research workers and by research engineers interested in the optimisation of processes. The methods of operational research, especially those involving the application of digital computers in simulation roles, are proving of considerable value in the application of the scientific method to industrial and applied scientific problems. This book provides an authoritative introduction to these topics.

CONTENTS

CHAPTER 1

INTRODUCTION

The subject-matter of this book stems from three origins and contains matters of interest to three groups of scientific workers.

The first and most respectable origin lies in the theory of mathematical statistics. Before and for some time after the founding of the Royal Statistical Society in 1834, the subject of statistics consisted of the collection and display in numerical and graphical form of facts and figures from the fields of economics, actuarial science and allied descriptive sciences. One of the most useful forms of display was the histogram or frequency chart and the transformation of statistics began when it was realised that the occurrence of such diagrams could be explained by invoking the theory of probability.

The idea of a probability distribution had been established as a useful concept by mathematicians and has been studied by Laplace and Gauss, to name two of the most celebrated mathematicians interested. The idea that frequency charts could be explained as a practical consequence of the laws of probability applied to everyday matters seized the imagination of the pioneers of mathematical statistics. Since a probability distribution is by its nature, in most instances, composed of an infinite number of items, and frequency charts by their nature are composed of a finite number of items, these latter had to be thought of as *samples* from an underlying theoretical probability distribution. The problem then raised itself at how to describe a probability distribution given only a sample from it. The mathematical difficulties of this seemed immense and such steps as were taken needed experimental verification to give the early workers confidence. Thus was born the sampling experiment. A close approximation to a probability distribution was created, samples were taken, combined and transformed in suitable ways and the resulting frequency chart of sampled values compared with the predictions of theory.

Although mathematical techniques have developed to levels of sophistication that would astonish the early workers, the value of sampling experiments in mathematical statistics still remains. The advent of automatic digital computers to perform the laborious calcula-

tions necessary has revitalised this as a possible approach to the solution of problems still beyond the reach of analysis.

The second origin lay in the demands of applied mathematicians for methods of solving problems involving partial differential equations. These equations constantly arise in the mathematical formulation of numerous physical processes. Analytical solutions were found for a wide range of formal problems but practical problems involving complex or irregular boundary conditions could not yield to theoretical attack. The corresponding situations had arisen in ordinary differential equations and had been overcome by developing numerical techniques for solving such equations. The nature of the problem for partial differential equations made an attack on these lines far more difficult.

A typical problem was the solution of the so-called diffusion equation, which arises, as its name implies, in the diffusion of gases as well as in the conduction of heat in a medium and many other physical systems. The characteristic of many of these systems was that the actual mechanism for the movement of the gas (or heat) involved a large number of particles behaving in a partly regular and partly irregular manner. Averaging over the particles enabled the random element to be eliminated and a deterministic description to be given.

Now the theory of probability had studied formal models closely allied to a system of particles moving in this partly regular, partly random manner and had developed mathematical techniques for dealing with these problems which were studied under the name of random walks. The mathematical analysis of these problems gave rise to partial differential equations of the same type as the diffusion problem. Thus was born the idea of solving experimentally the diffusion and allied equations by random walks. The idea lay dormant for many years and was resurrected by Von Neumann and Ulam under the stress of the technological demands of the Second World War.

These men conceived of the extension of the principle to seek solutions to difficult mathematical problems arising from deterministic problems by finding analogous problems in probability leading in their analysis to formally identical mathematical equations and then solving the probability problems practically by sampling experiments. By this time, the ultimate stochastic nature of the physical phenomena often present in the original problem was often forgotten or completely ignored.

One of the simplest and most powerful applications of this idea, which they christened 'Monte Carlo', was the evaluation of a multi-dimensional

integral. Consider the simplest case of evaluating the area of a bounded region. Surround the region with a square, scaling to make its side of unit length. Take a point in the area at random. The probability that this lies in the region of area A is simply A. Take a large number of points at random; the proportion of these lying in the area is an estimate of A.

The whole idea can be generalised to higher dimensions and the result remains true. The great advantage of this method lies in the fact that for higher dimensions, conventional numerical techniques will require an enormous number of points to get any answer at all. The Monte Carlo method can get an answer with any number of points although of course the accuracy falls off as the number decreases.

In fact, the accuracy is proportional to $\sqrt{n}$, n being the number of points. Conventional methods give an accuracy proportional to $n^{1/d}$ where d is the number of dimensions.

The third origin lay in the new science of Operational Research. This set itself the tasks of applying scientific method in everyday life—to military problems during the last war and to problems of industry after it. In common with most sciences, it developed by building models of the systems it studied and using these models to give insight and sometimes quantitative information about these systems.

The outstanding difference between the subject-matter of conventional science and operational research lay in the greater variability of the phenomena studied. It was vital to bring regularity back into the description of these phenomena—to find a usable description of the variability. This was achieved, as had been done for economics and actuarial science in the previous century, by using probability theory.

The models studied were primarily probabilistic and the theoretical difficulties of these models were immense. A special branch of probability theory arose under the title of Queue Theory to deal with some of these problems, but the irregular nature of the boundary conditions and restrictions in the real problems could rarely be incorporated into the models that mathematical methods could solve.

Once again the scientist turned to an experimental technique. Now the problem was phrased in a language description of the real world involving queues, stocks, machines and operating instructions. The subject of simulation, as it was called, took on an identity of its own.

Yet this was nothing new. To a statistician the problem was nothing more than to find the sampling distribution of an intricately and irregu-

larly defined statistic and because of the intricate nature of the definition, to do this by a sampling procedure.

The common element in all these approaches to our subject is probability theory. In these circumstances, it is hardly necessary to stress that an understanding of at least the elements of the theory are essential to an understanding of the principles and practice of sampling.

Any attempt to give a self-contained account of the theory necessary as part of the book would be self-defeating. It could be nothing more than a few formal definitions and readers ignorant of the subject might be tempted to read further with only such a brief introduction. Therefore it is assumed that the reader has the necessary grounding in elementary probability theory.

Likewise, the elements of the theory of statistics are assumed. The commonest distributions such as Normal, Student's, χ^2, binomial, exponential, are all discussed without a long introduction and the χ^2-test of goodness of fit is not explained, as this must surely be well known to all scientific workers.

Since the basic process involved is the drawing of samples from different distributions, this receives first study. The problem is reduced to drawing a sample from a uniform distribution (or drawing an integer from a range) 'at random'.

The methods of solving *this* problem by both machinery and mathematical methods are discussed.

This is followed by an account of the problems in determining a frequency distribution by sampling.

The possibility of using sampling methods as *practical* methods of acquiring knowledge rests on the possibility of taking really large samples to reduce the effect of the inherent variability. This in turn is possible only because of the existence of automatic digital computers.

It is assumed that one of these powerful devices is available and this raises the problem of describing how to use these machines for sampling. Again, this book is not the place to attempt to give details of the problems of instructing any particular machine to perform the required calculation.

However, the idea of a flow diagram as a means of describing the calculating procedure in a form suitable for transcription into a machine code is explained and forms a powerful means of reducing the volume of description of computing procedures.

Probability distributions are described by means of a set of parameters and the central problem of sampling is to specify the value of

such parameters given a sample from the distribution. This is the problem of estimation, which receives full treatment. An estimate must be as accurate as possible, and methods of sampling and estimation that increase the accuracy of the estimate without a corresponding increase in the volume of computing are described.

We are then ready to tackle the problems of simulation. First simple queues are dealt with, gradually increasing in difficulty until a general case is dealt with.

More complex systems are then treated and a general technique evolved for dealing with any system.

Finally the question of the design of such large-scale experiments is considered.

CHAPTER 2

SAMPLING FROM A DISTRIBUTION

2.1 Introduction

We have seen in the preceding chapter that the basic problem for simulation consists in sampling from statistical distributions, and this chapter will describe a variety of methods of performing this basic operation.

There are two main types of distributions from which samples are required: those in which the statistical variable takes a continuous range of values giving rise to a continuous probability density function, and those in which the statistical variable can take a discrete number of values. The normal distribution is typical of the first class and the binomial distribution (the number of successes in a fixed number of similar independent trials) is an example of the latter.

However, for practical purposes, we must work with approximations to continuous variables and have to be satisfied with samples from a grouped distribution represented by a histogram. If the different groups are given names, then the problem reduces to selecting a name for the appropriate frequency and consequently there is no distinction in practice between the two types of distribution.

Although it is possible to discuss the meaning of probability at great length, the earlier remarks indicate that our purposes are met by a frequency concept. Thus our problem is to devise a process of selection from the distribution so that the results of the repetition of this process will give rise to a frequency distribution of sampled values that matches the frequency distribution required in our simulation. It is possible for there to be quite intricate probability rules relating successive sample values, but the most important case that will be considered in detail first is where these samples are independent of one another in the sense that the probability distribution for any given item from the sample is independent of the preceding sample values.

To effect the random selection, some source of randomness is required, and to give a general solution to the problem we want to use a common source for all types of distribution. The most appropriate source of such randomness is from random numbers. These may be

supplied either by tables, specially constructed by means of randomising machines, or by the use of mathematical devices to generate long sequences of numbers that have similar properties to random numbers, the so-called pseudo-random numbers. These techniques will each be pursued in later chapters.

We consider first the elementary methods of selecting from a frequency distribution and show in a series of steps how this can be improved and mechanised.

2.2 The 'Top Hat' Method

The obvious method to obtain a sample from a frequency distribution has often been used in early historical examples of simulation. It consists of representing each of the possible values of the statistical variable by means of a set of discs. The number of discs given any particular value is proportional to the frequency with which it occurs in the frequency distribution. The total number of discs used is limited in an upward direction by the problems of the physical space occupied by the discs and in a lower direction by the accuracy with which the frequencies must be approximated. The collection of discs is well shuffled in a hat or other receptacle and a disc drawn at random. The values on successive discs after drawing (each being replaced after it has been read) constitute the sample.

As an example, suppose we wish to draw a sample from a normal distribution with mean 18 units and standard deviation 4 units. The first step is to decide on the width of grouping that will be tolerable. This is settled by the application and we suppose a class width of 2 units is chosen, and the class boundaries are taken symmetrically about the mean.

The next step is to determine the probability of this random variable falling in the intervals -1 to 1, 1 to 3, 3 to 5, ..., 15 to 17, ... and these are obtained from a table of the normal distribution function. This is given in Table 1.

If we decide to represent the frequencies to three decimal places, we will require 1,000 discs. The frequency of negative values is negligible. Thus we number 28 discs with the number 10 representing the class interval from 9–11, 2 with 6 representing the interval 5–7, etc. In practice, the rounding has to be adjusted to give the sum of the frequencies exactly 1,000 or a few discs rejected. From the table given one disc is rejected.

This example raises a problem of dealing with the tails of distribu-

TABLE 1

Class	*Frequency*	*No. of discs* (*total* 1,000)
−1	0·0000	0
1– 3	0·0001	0
3– 5	0·0005	1
5– 7	0·0024	2
7– 9	0·0092	9
9–11	0·0279	28
11–13	0·0655	65
13–15	0·1210	121
15–17	0·1747	175
17–19	0·1974	197
19–21	0·1747	175
21–23	0·1210	121
23–25	0·0655	65
25–27	0·0279	28
27–29	0·0092	9
29–31	0·0024	2
31–33	0·0005	1
33–35	0·0001	0
≧35	0·0000	0
		999

tions which are unbounded. In order to represent such distributions correctly, we need an infinite number of discs so that the extreme tails with their correspondingly extreme small probabilities can be represented at all. This lays a practical limit on the length of tail approximated.

However, this difficulty can be overcome in the following way. Let the top group of the distribution represent the frequency with which an item should lie beyond the last bounded group. If a disc representing this unbounded group is obtained, then a further sample is taken from a distribution that gives the proportion of the items having a given value conditional upon the item being in the upper tail. Thus to accommodate the possibility of sample values in the range greater than 33 the following Table 2 of relative frequencies is constructed from more accurate tables of the normal integral.

The primary labelling of discs is adjusted to allow for the extra unbounded class, and a secondary labelling made according to the second distribution. If the first sampling gives rise to an item belonging to the upper tail, another disc is selected at random and the second value

TABLE 2

Class	*Frequency*	*Relative frequency*	
33–35	0·0000677	0·8635	864
35–37	0·0000097	0·1237	124
37–39	0·0000009	0·0115	11
≧39	0·0000001	0·0013	1
	0·0000784	1·0000	1,000

on that is used as a direct sampled value. The lower tail can be dealt with similarly. The tails of the conditional distributions can be sampled in the same way and there is no limit to the accuracy obtainable.

There are two main objections to this elementary method of sampling. The first is the very practical one that the whole sampling process is rather laborious and the rate of sampling is limited by the rate that discs can be shuffled and selected. As we shall see, it is essential for large-scale simulations to mechanise the arithmetic processes involved, and this implies for their effective execution that the process of random selection should also be mechanised.

The second and more serious objection is that in practice the shuffling process that can be applied to these discs is not adequate and many trials have been made which show that samples drawn in this way are not in fact random. To illustrate the point, the following experiment was performed.

TABLE 3

Number		$(x - \bar{x})^2$
1	51	1
2	46	16
3	38	144
4	57	49
5	60	100
6	50	0
7	58	64
8	41	81
9	51	1
10	48	4
—	500	460

A set of 1,000 discs labelled with numbers from 1 to 10 (100 of each) were prepared and sampled in the way proposed. The result of 500 drawings is given in Table 3.

Column 1 gives the actual frequency of occurrences of the different symbols. The expected number of occurrences of each class is 50. Applying a χ^2 test gives $\chi^2 = 9{\cdot}2$ which is significant at the 1% level. It is clear that the samples of 4, 5, 7 drawn from this distribution were markedly higher than expectation, whilst samples of 3, 8 were below expectation.

2.3 A Mechanised Version of the 'Top Hat' Method

This can best be understood if we consider how the selection of the discs in the 'Top Hat' method could be made reliably random. Supposing in addition to the variate values inscribed in the discs, the discs are numbered serially. Our requirement is that on each drawing any one of the discs could be equally likely to be selected, and this implies that the serial number on the selected disc is equally likely to be any of those present. Thus the random selection can be made secure by ensuring the serial number of the disc drawing is a random number selected from a reputable source. The validity of the selection process now depends entirely on the validity of the source of random numbers used in the sampling process.

This, of course, does not overcome the objection to the clumsy nature of the process, but gives the key to mechanised techniques for doing it. In effect, the collection of discs represent a function. This function has, as its independent variable, the serial number, and as its value, the variable inscribed on the disc. The whole apparatus of discs can be discarded and the process applied by using a table of values of random variables, each associated with a serial number. Since the serial numbers will be selected at random, there is no restraint on the arrangements of the values of the variable on the discs, and the simplest arrangement is to use the lowest serial numbers for the lowest variable value, the next serial numbers for the next lowest variable and so on, so that the variable values as a function of a serial number are monotonic. The discs, or items, can be thought of as being used to build up a histogram of discs as shown in Figure 1.

In this figure, each cell has the serial number in the top left-hand corner and the variate value in its body. The variate values in any given column are thus constant. Thus, using this example, if our source of random numbers gives the random numbers 34, 18 and 41 respectively,

the three successive samples chosen from this distribution will be 6, 4 and 7.

The function defined in this manner is a rather flat one in that the variate values only occasionally change value with the serial number, and it is possible to compress the information considerably by listing

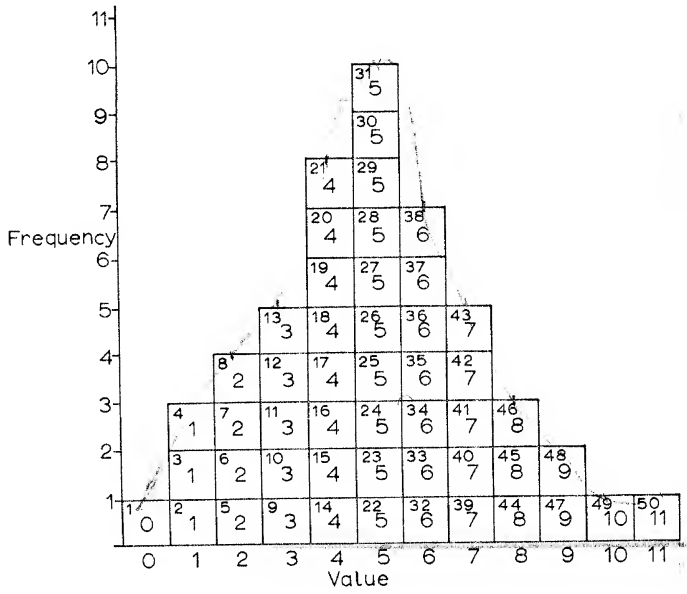


Figure 1

only the critical points in a table. These are the serial numbers of those items at which the variate value has changed. Thus the histogram can be turned into the following Table 4:

TABLE 4

Value	0	1	2	3	4	5	6	7	8	9	10	11
First serial No.	1	2	5	9	14	22	32	39	44	47	49	50

If the random number selected is 33, since this lies between the critical numbers 32 and 39, the value of the variate sample is taken as 6. This compression of the tabular function is available even for hand simulations, but is of greater importance if it is intended to use a digital computer for the selection process as it saves a considerable amount of storage space.

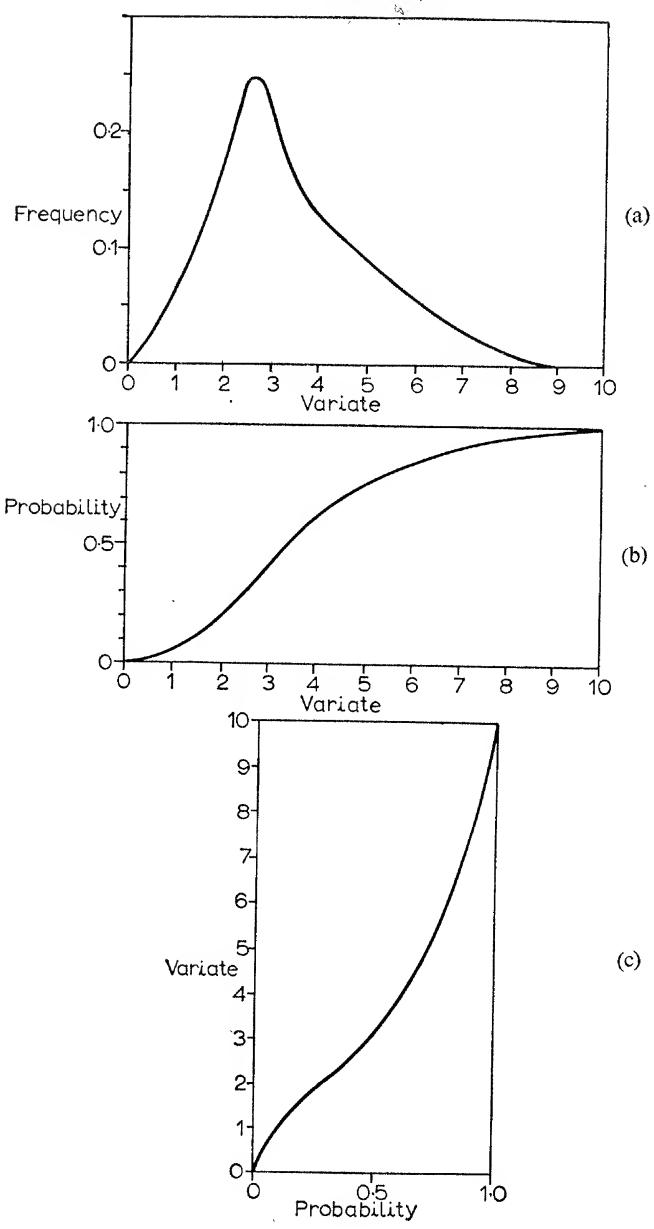

0·2
Frequency
0·1
0
0 1 2 3 4 5 6 7 8 9 10
Variate
(a)
1·0
Probability
0·5
0
0 1 2 3 4 5 6 7 8 9 10
Variate
(b)
10
9
8
7
6
5
Variate
4
3
2
1
0
0 0·5 1·0
Probability
(c)

Figure 2

This tabular function has an interesting probabilistic intepretation. The distribution of a variable can be described by a density function in the continuous case or by a set of frequencies in the discrete case, but both cases (and other more complicated ones) can be described by means of the cumulative distribution function which specifies the probability of obtaining the given value or less from a sampling distribution. The tabular function introduced above is the inverse of this, and this is illustrated in Figures 2(a), 2(b) and 2(c).

Figure 2(a) gives a typical density function, 2(b) gives its cumulative distribution function and 2(c) its inverse. The table look-up technique introduced before consists of choosing a point on the independent axis of the inverse cumulative distribution and choosing the corresponding dependent value as the sample value. It is very easy to see that this process will reproduce the required distribution as follows:

Let r be the random number in the range 0–999 say. Then the distribution of $u = \dfrac{r}{1000}$ is approximately uniform over interval 0–1.

Then
$$p(u) = \begin{cases} 0 & 0 & u \leqq 0 \\ 1 & P(u) = u & 0 \leqq u \leqq 1 \\ 0 & 1 & u \geqq 1 \end{cases}$$

where $p(u)$ and $P(u)$ are the probability density function, and cumulative distribution function respectively of u.

If x is to have a p.d.f. $f(x)$ with c.d.f. $F(x)$ then let the inverse of F be Φ

$$y = F(x) \qquad x = \Phi(y)$$

Put
$$x = \Phi(u)$$

Then

$$\text{Prob}(x \leqq X) = \text{Prob}(F(x) \leqq F(X)) = \text{Prob}(u \leqq F(X)) = F(X)$$

Thus the c.d.f. of x is $F(x)$ as required.

Viewed in this light, it is seen that the problem of sampling from any distribution is that of transforming a random number representing the uniform random variable in the range of 0–1 by means of the inverse cumulative distribution function. The techniques considered so far define this function by means of a table, but there are other methods available for the representation of a function, and many of these can be applied in this case. The appropriate methods consist of:

(i) Analytic inversion of the cumulative distribution function and the calculation of the value of this function for the value of a selected uniform random variable.
(ii) Numerical inverse interpolation in the distribution function determined analytically.
(iii) A process of numerical inverse interpolation in a numerical approximation to the cumulative distribution function.
(iv) The numerical approximation to the inverse cumulative distribution function itself.

Each of these methods has its merits in various circumstances, and we will consider them in turn.

(i) *Analytic Inversion.*

The probability density functions of theoretically derived distributions are often complex in form and the possibility of analytic inversion of their integrals is usually difficult. However, there are some simple but quite important cases in which this is possible. For example, a commonly arising distribution is the negative exponential, which has p.d.f. of the form

$$p(x) = e^{-x}$$

and c.d.f.

$$P(x) = 1-e^{-x}$$

The inverse of this is simply

$$x = -\log_e (1-P)$$

Since the distribution of $1-P$ is the uniform over 0–1, if the distribution of P is uniform over 0–1 we have the simple rule:

To find an exponentially distributed variable to take the negative of the logarithm of a uniform random variable.

The convolution of the two uniform variables gives a triangular distribution defined by

$$p(x) = \begin{cases} 0 \\ x \\ 2-x \\ 0 \end{cases} \qquad P(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2}x^2 & 0 \leqq x \leqq 1 \\ 2x-\frac{1}{2}x^2-1 & 1 \leqq x \leqq 2 \\ 1 & x > 2 \end{cases}$$

$P(x)$ is a function which can be inverted by the following rule:

$$P \leqq \tfrac{1}{2} \qquad x = \sqrt{2P}$$

$$\tfrac{1}{2}P \leqq 1 \qquad x = 1-\sqrt{2(1-P)}$$

However, as the convolution is merely the distribution of the sum of two random variables, in this case it may be more economic merely to sample two uniform random variables and add them.

A third distribution which can be treated in this way is the Cauchy distribution with probability density function:

$$p(x) = \frac{1}{\pi}\,\frac{1}{1+x^2}$$

This gives the c.d.f. as

$$P(x) = \tfrac{1}{2}+\frac{1}{\pi}\tan^{-1} x$$

and thus the inverse is given by

$$x = \pi \tan (P-\tfrac{1}{2})$$

(ii) *Numerical inverse interpolation* in the analytic function.

There are cases in which it is possible to integrate the probability density function and represent the c.d.f. analytically, but the inversion is not possible analytically. In this case it is possible to proceed by generating the analytic function for a range of x in the neighbourhood of the required value and then using the random number selected as an interpolate for the required value. This is a tedious process for hand techniques, and it would only be considered on digital computers. In practice, the number of c.d.f.s that can be calculated by direct substitution in an analytic formula are rather small, and this method is usually replaced by

(iii) *Numerical inverse interpolation* in a numerical approximation to the c.d.f. In this case, an approximation to the c.d.f. is found by an algebraic formula, values are generated and interpolated. However, most effective numerical approximations to the c.d.f.s are quite arbitrary and empirical and under these circumstances it seems advisable to work directly with an approximation to the inverse of the c.d.f. and so avoid the necessity for the inverse interpolation. This is method (iv) which will be considered in some detail.

(iv) *Approximation to Inverse c.d.f.*

The general shape of the function required is given in Figure 2(c) and this has the properties that

$$F \to \infty \quad \text{as} \quad x \to 0 \quad , \quad F(x) \to +\infty \quad \text{as} \quad x \to 1$$

(where x is now written for the independent variable P) and so the function is likely to satisfy a differential equation of the form

$$F'(x) = \frac{a}{x} + \frac{b}{1-x} + \phi(x)$$

where $\phi(x)$ is a regular function (i.e. has no singularities) and a and b are constants or bounded functions for x in the range [0, 1]. Taking a and b as constants gives

$$F(x) = a \log x + b \log (1-x) + \Psi(x)$$

where Ψ is another regular function.

In practice it is found that this form is not very satisfactory and certain arbitrary adjustments have to be made to it to give better fit to a wider class of functions. The form finally found to be most useful is

$$F(x) = a + bx + cx^2 + \alpha(1-x)^2 \log x + \beta x^2 \log (1-x) \qquad (1)$$

This satisfies the requirements set out before and uses a simple polynomial for $\Psi(x)$ and replaces the constants a and b by functions to give better fit in the tails. In practice this formula is difficult to use because of the logarithmic function. If it is being used in manual calculations, then tables of logarithms can be used, but the merits of this method lie in the possibility of its automation and in this case, the calculation of the logarithm is a comparatively slow process and thus an approximation to the logarithm is necessary. It is clear that the base of the logarithm is immaterial, as this merely alters the constants α and β by a conversion factor, so we use the base 2 and use the crude approximation

$$\log_2 x \simeq \log_2 X.2^{-d} \simeq -d + (1-X)$$

where

$$x = X.2^{-d} \quad , \quad \tfrac{1}{2} \leqq X < 1$$

The advantage of this method is that we may now treat all distributions in the same way whether they be analytically derived or arising from practical data. In both cases, a method of determining the constants a, b, c, α and β is required which we shall consider later.

As an example of this method, it has been found that for a normal distribution with mean μ and standard deviation σ the following formula gives an adequate approximation.

$$a = (16384\mu - 13452{\cdot}960\sigma) \times 2^{-11}$$
$$b = (26953{\cdot}865\sigma) \times 2^{-11}$$
$$c = 0$$
$$\alpha = (-3772{\cdot}769\sigma) \times 2^{-11}$$
$$\beta = (3772{\cdot}769\sigma) \times 2^{-11}$$

The difference between the true and approximate cumulative distributions is shown in Figure 3.

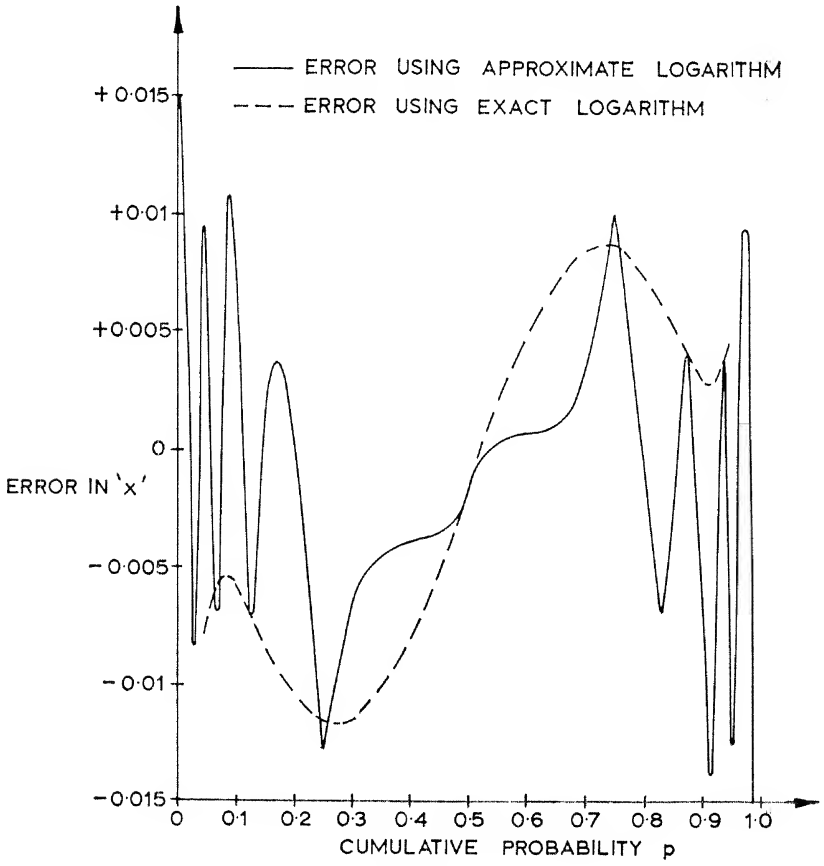


Figure 3

It will be seen that the approximate function has a discontinuous derivative which is caused by the approximation to the logarithm that is used. The approximation using an exact logarithm is also plotted, which shows that the error introduced by the logarithmic approximation is less than the error of the main approximation, and that the maximum total error is very small.

The approximation works successfully in the case of unimodal distributions, but is not successful for J-shaped distributions or multimodal distributions.

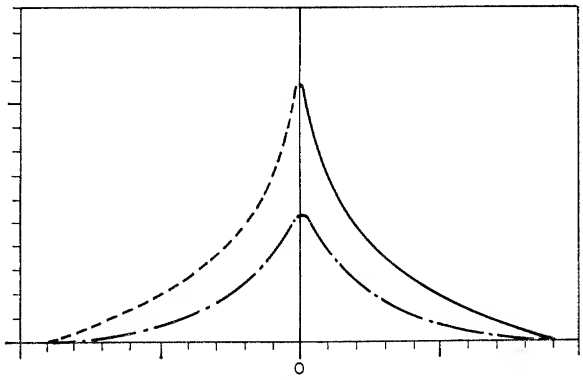


Figure 4

In the case of J-shaped distributions, a simple artifice will enable the technique to be used. Assume the distribution has an origin at zero and can be represented by the heavy line in Figure 4.

We reflect this distribution in the y axis (dotted line) and rescale to give the chain-dotted distribution. This is a symmetric unimodal distribution and the transformation may be used. By taking the absolute value of a sample from this distribution a sample from the original distribution is obtained.

If the distribution is multimodal, it can be represented as a mixture of several distributions each unimodal, and expressed by a suitable approximation:

$$F(x) = \sum_i p_i f_i(x)$$

The interpretation of this is that the observation has come with probability p_i from a distribution with p.d.f. $f_i(x)$. Techniques for the determination of the component distributions have been developed, using the

method of moments, and more recently, techniques using the method of maximum likelihood to estimate the parameters of the system using a digital computer have been used.

A random variable is used to select which distribution shall be sampled. The distribution chosen is then sampled and the resulting variable will be a proper selection from the multimodal distribution.

Thus to represent the distribution

$$F(x) = \sum_{i=1}^{3} p_i N(\mu_i, \sigma_i) \quad (\sum p_i = 1)$$

two uniform random variables R_1, R_2 are chosen. A random normal deviate $X(N(0, 1))$ is selected using the variate R_1 and the approximation described earlier.

The variate R_2 is used to select the population sampled.

$$\left.\begin{array}{ll} \text{If} & R_2 < p_1 \\ \text{If} & p_1 \leqq R_2 < p_1 + p_2 \\ \text{If} & R_2 \geqq p_1 + p_2 \end{array}\right\} \quad \text{take} \quad \left\{\begin{array}{l} x = \mu_1 + X\sigma_1 \\ x = \mu_2 + X\sigma_2 \\ x = \mu_3 + X\sigma_3 \end{array}\right.$$

The technique used for obtaining the coefficients a, b, c, α, β of equation (1) should be a general one which is equally applicable if the distribution in question is given theoretically or empirically. We shall deal first with the case of an empirical distribution. From the data it is possible to construct a table giving for the various cumulative frequencies the corresponding variate value. We require to pass the curve of 'best fit' through this series of points and a possible method to use is the method of least squares. There is no theoretical justification for this process as the observations to which this method is applied are not independent and do not have a homogeneous variance. Nevertheless, it is found that in practice this method gives adequate values for the coefficients a, b, c, α, β. It is always advisable to verify the validity of the fitted function by calculating the function for a series of values of x and deriving an approximation to the p.d.f. by inverse interpolation and numerical interpolation or by graphical means.

There are many applications in which the exact nature of the distribution is not known, because it is derived from empirical data and in this case it is only necessary to check the robustness of the answers from the sampling experiments performed against changes in the basic distributions used. If an automatic computer is used to do the sampling

processes, this is a possibility, but otherwise it would be prohibitively heavy in labour.

For a theoretical distribution the percentiles are obtained and a similar fit by least squares is used on this as if it were data from an empirical distribution.

It is at this point that the distinction between discrete and continuous distributions is brought to the fore again. For a discrete distribution, the jumps in the distribution function are real and we require our function to pass through the saltuses. The values taken are then rounded down. For a continuous distribution, these jumps have been introduced artificially by the grouping process, and it is best for the function to pass through the mid-points of the saltuses. In the case of a theoretical distribution where the P-axis increments can be arranged constant, we need only fit to the left-hand end points and then adjust the constant a in the formula by half the percentile interval used.

If it is required to determine the mean and variance of the fitted distribution, these can be found from the coefficients a, b, c, α, β. If the mean and variance are μ and σ^2 respectively then

$$\mu = \int xp(x)\,dx = \int x\,dP(x)$$

$$\sigma^2 = \int x^2\,dP(x) - \mu^2$$

But

$$x = a + bP + cP^2 + \alpha(1-P)^2 \log P + \beta P^2 \log P$$

$$\mu = \int_0^1 (a + bP + cP^2 + \alpha(1-P)^2 \log P + \beta P^2 \log (1-P)\,dP$$

$$= a + \frac{b}{2} + \frac{c}{3} + (\alpha+\beta)\int_0^1 (1-P)^2 \log P\,dP$$

The integral can be evaluated numerically from the approximation to $\log P$ used and we obtain

$$\int_0^1 (1-P)^2 \log P\,dP \doteq 0{\cdot}904$$

Then

$$\mu \doteq a + \frac{b}{2} + \frac{c}{3} + 0{\cdot}904(\alpha+\beta)$$

Similarly we can obtain σ^2 as a quadratic function of the parameters b, c, α, β. The parameter a which controls the location does not appear, of course.

$$\sigma^2 = [b, c, \alpha, \beta]\Omega[b, c, \alpha, \beta]'$$

where

$$\Omega = \begin{bmatrix} +0{\cdot}083333 & +0{\cdot}083333 & -0{\cdot}312104 & +0{\cdot}312104 \\ +0{\cdot}083333 & +0{\cdot}088888 & -0{\cdot}365733 & +0{\cdot}258474 \\ -0{\cdot}312104 & -0{\cdot}365733 & +1{\cdot}973261 & +0 \\ +0{\cdot}312104 & +0{\cdot}258474 & +0 & +1{\cdot}973261 \end{bmatrix}$$

For a symmetrical distribution, with mode m, we must have for each x

$$x = F(P)$$

$$m-x = F(1-P)$$

$$\text{i.e.} \quad F(P)+F(1-P) = m$$

$$2a+b+c)-2cP+2cP^2+(\alpha+\beta)\{(1-P)^2 \log P+P^2 \log (1-P)\} = m$$

From this the necessary conditions for symmetry are

$$\alpha = -\beta$$
$$c = 0$$

when the mode is given by

$$m = 2a+b$$

The median is given by

$$x = F(\tfrac{1}{2}) = a+\tfrac{1}{2}b+\tfrac{1}{4}c+(\alpha+\beta)\tfrac{1}{4} \log \tfrac{1}{2}$$
$$= a+\tfrac{1}{2}b+\tfrac{1}{4}c-\tfrac{1}{4}(\alpha+\beta)$$

This confirms that the mean and median only coincide for symmetric distributions. However, it may happen that a better fit is obtained by relaxing the condition of symmetry. This occurs, for example, with the approximation quoted for the normal distribution.

If a quantile measure of dispersion is used the semi inter-quantile range R can be easily calculated

$$R = F(\tfrac{3}{4})-F(\tfrac{1}{4})$$
$$= \tfrac{1}{2}(b+c)+\tfrac{1}{16}(\beta+\alpha)(9 \log \tfrac{1}{4}-\log \tfrac{3}{4})$$

An alternative fitting procedure uses the method of minimum χ^2. For any choice of parameters a, b, c, α, β, the probabilities P_i falling in a range can be obtained by inverse interpolation and the actual frequencies (in case of an empirical distribution) or the theoretical probability (in case of a theoretical distribution) compared by compiling a χ^2 statistic. The parameters can now be estimated to minimise this quantity.

This procedure involves the solution of non-linear equations and can only be achieved by an iterative process. A method is that of steepest descent.

This method involves an arbitrary element in the classes used to sub-divide the range and different answers are obtained from different divisions. However, all possible solutions will approximately reproduce the distribution.

The arbitrary nature of these techniques is merely a reflection of the arbitrary nature of the fitting problem. With a limited number of parameters the fit can only be improved at one region of the distribution at the expense of another region.

A very simple approach is to specify 5 quantiles of the distribution. Suppose the quantiles $Q_1, Q_2, \ldots, Q_5$ are to be at $x_1, x_2, \ldots, x_5$. Then the equations

$$x_i = a + bQ_i + cQ_i^2 + \alpha Q_i^2 L(1-Q_i) + \beta(1-Q_i)^2 L(Q_i) \qquad L = 1, 2, \ldots, 5$$

are linear in a, b, c, α, β and on solving these a fit is obtained which gives the desired values in the neighbourhood of $x_1, x_2, \ldots, x_5$.

For a symmetric distribution, the necessary restraint must reduce the independent qualities that can be fitted to 3. The good fit for the normal distribution implies that for markedly non-normal symmetric distributions the fit cannot be so good. This can be remedied by replacing the polynomial terms by a higher order polynomial and at the expense of more calculation during sampling any accuracy can be attained.

The logarithmic terms could be dispensed with but their inclusion greatly reduces the order of polynomial required.

2.4 The Method of Mixtures

The technique used to sample from a multimodal distribution can often give fast sampling techniques for unimodal distributions. It is particularly useful if very accurate sampling is required.

The distribution is represented as the mixture of a set of rectangular distributions and a residual. This is best explained by reference to Figure 5.

Figure 5(a) gives the distribution to be sampled, partitioned into 3 distributions. Figure 5(b) is a rough histogram which includes about

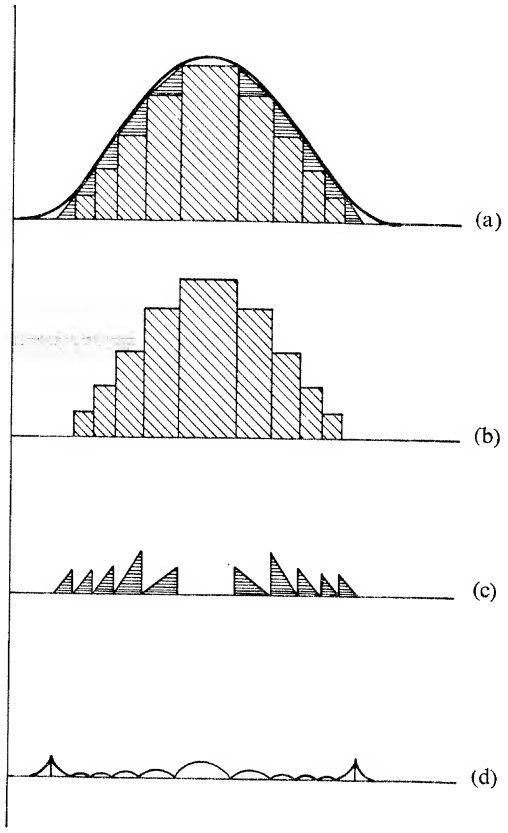


Figure 5

90% of the density. The set of triangular distributions of Figure 5(c) account for all but about 0·2% of the remaining density. This latter has the distribution of Figure 5(d).

If the probability area under these functions are p_1, p_2, p_3, a random number is used to select the distribution to sample.

If component (*a*) is selected, a second random number is used to

select a rectangular component and then a third random number is scaled and its origin adjusted to give the required value.

If component (*b*) is selected, a triangle is selected with a second random number and a third random number is converted into the required value by a square root transformation.

If component (*c*) is selected, a more complicated procedure will be necessary, but this will rarely occur and the average time taking a sample is quite small.

For theoretical distributions, the class boundaries of the histogram can often be chosen, so that they or the class areas are easy to calculate and need not be stored. Otherwise, the method can be quite costly in storage space.

2.5 The Rejection Procedure

It is often possible to calculate the p.d.f. of a variable fairly easily, but difficult to evaluate its integral or the inverse. A technique called the rejection technique has been developed to deal with this situation.

Consider a bounded p.d.f., i.e. one for which there exist values a, b, M such that

$$p(x) = \begin{cases} 0 & x < a \\ f(x) & a \leqq x \leqq b \qquad 0 \leqq f(x) \leqq M \\ 0 & b > 0 \end{cases}$$

For any x in range $[a, b]$ calculate $y = \dfrac{f(x)}{M}$; the probability that a uniform random variable R will not exceed y is precisely y. If the value x is chosen at random on the range $[a, b]$ and then rejected if $R > y$, the p.d.f. of accepted x's will be $f(x)$.

This follows since the probability of a given pair x, R giving an accepted value x is $y\,dx$ and the probability of being accepted for any x is

$$\int_a^b y\,dx = \frac{1}{M}\int_a^b f(x)\,dx = \frac{1}{M} \quad \text{since } f(x) \text{ is a p.d.f.}$$

Thus the conditional p.d.f. of an accepted pair having a fixed x-value is $My = f(x)$.

The number of trials before a successful pair is found is a random variable and has a distribution

$$p(n) = E(1-E)^{n-1} \quad \text{where} \quad E = \frac{1}{M}$$

This has a mean value

$$\sum_{n=1}^{\infty} nE(1-E)^{n-1} = -E\frac{d}{dE}\left\{\sum_{n=1}^{\infty}(1-E)^n\right\}$$

$$= -E\frac{d}{dE}\left(\frac{1-E}{E}\right) = \frac{1}{E}$$

The technique will not work without modification if the distribution required is unbounded.

The modification consists of using an unbounded random variable to make the primary selection, this being obtained by a transformation of a uniform variable as previously described.

The rejection criterion can also be generalised and the complete procedure is as follows:

Put $$p(x) = kQ(x)r(x)$$

where $r(x)$ is a p.d.f., $Q(x)$ is c.d.f. and k is a normalising constant. Select x from the p.d.f. $r(x)$ and y from the c.d.f. $Q(y)$. Accept the pair if $x < y$ and use x as the required variable.

The probability of choosing the x is $r(x)\,dx$ and of accepting it is $Q(x)$.

The probability that the pair is acceptable is

$$\int Q(x)P(x)\,dx = \frac{1}{k}$$

so the conditional probability of an accepted pair having a fixed x-value is $kQ(x)P(x)\,dx$ as required.

It is often possible to effect the test $y \leqq x$ without calculating $Q(x)$.

If $f(\xi)$ is any increasing monotonic function with an inverse ϕ the test $y < x$ is entirely equivalent to $f(y) < f(x)$. The c.d.f. of $z = f(y)$ is given by $Q(\phi(z)) = T(z)$, say.

If f and hence ϕ is chosen properly, the function $T(z)$ may be simple and the procedure becomes:

Select x from $R(x)$ and z from $T(z)$ and accept the value x if and only if $z < f(x)$.

Since $$Q(x) = \frac{p(x)}{r(x)}\,\frac{1}{k}$$

$$0 \leqq \frac{p(x)}{r(x)} \leqq k$$

By an argument similar to that for the simple case, the mean sample size is k and so to economise on sampling, $r(x)$ should be chosen so that

$$\max_x \frac{p(x)}{r(x)} \quad \text{is minimised.}$$

In practice this means choosing $r(x)$ to be similar in shape to $p(x)$. This is in accord with the dictates of common sense since then the amount of rejection to convert the primary selection probability for x from $r(x)$ to $p(x)$ will be small.

A similar technique can be used writing

$$p(x) = k[1-Q(x)]r(x)$$

and only accept a pair if $y > x$. The argument and extension are exactly similar to the preceding case.

To illustrate the technique, consider the normal distribution. Start with the distribution of the absolute value z of a normal variate of zero mean and unit standard deviation. This can be converted to a normal variable by assigning a sign at random.

$$p(z) = \sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}z^2} \qquad 0 \leqq z < \infty$$

$$= \sqrt{\frac{2e}{\pi}}\, e^{-\frac{1}{2}(z-1)^2} e^{-z}$$

Now the p.d.f. of a negative exponential distribution is $r(x) = e^{-x}$ and its c.d.f. is $Q(x) = 1-e^{-x}$. Then we may write

$$p(z) = \sqrt{\frac{2e}{\pi}}\, \{1-Q(y)\}r(z)$$

where $y = \frac{1}{2}(z-1)^2$.

Thus if y and z are selected from a negative exponential distribution and the pair only accepted if $y \geqq \frac{1}{2}(z-1)^2$, z will have a normal distribution if a random sign is allotted.

This technique is most useful in dealing with empirical distributions that can be roughly described by a standard easily derived distribution and need 'correction' by a rejection process. The details of this application depend on the computing sources available and will not be discussed further.

The rejection technique can be amalgamated with the composition technique and the required p.d.f. represented as

$$p(x) = \Sigma\alpha_i p_i(x) = \Sigma\alpha_i Q_i(x) r_i(x)$$

where the $Q_i(x)$ are c.d.f.s, $r_i(x)$ are p.d.f.s and α_i are constants scaled to ensure that $\int p(x)\,dx = 1$.

Select a value of i with probability $\alpha_i/\Sigma\alpha_i$ and then sample from $p_i(x)$ by the rejection technique.

The mixed scheme has been used by Batchelor to give a very fast procedure for the normal distribution. As in the previous example the modulus is first found and then a sign allotted at random.

Write:

$$\sqrt{\frac{2}{\pi}}\, e^{-\frac{1}{2}x^2} = \sqrt{\frac{2}{\pi}}[e^{-\frac{1}{2}x^2}\, r(x) + \tfrac{1}{2}\, e^{-\frac{1}{2}(x-2)^2}\, s(x)]$$

when

$$r(x) = \begin{cases} 1 & 0 \leqq x \leqq 1 \\ 0 & x > 1 \end{cases} \qquad s(x) = \begin{cases} 0 & , \quad 0 \leqq x \leqq 1 \\ 2e^{-2(x-1)}, & x > 1 \end{cases}$$

i.e. $r(x)$ is the p.d.f. of a uniform random variable and $s(x)$ is that of a variable formed by adding 1 to an exponential variate of mean $\frac{1}{2}$.

Sample from $r(x)$ with probability $\frac{2}{3}$ and from $s(x)$ with probability $\frac{1}{3}$, i.e. select a random number u and

(a) sample from $r(x)$ if $u - \frac{2}{3} < 0$
(b) sample from $s(x)$ if $u - \frac{2}{3} > 0$

In the former case u has a uniform distribution in $[0, \frac{2}{3}]$ and in the latter case $u - \frac{2}{3}$ has a uniform distribution in $[\frac{2}{3}, 1]$.

In case (a) take $x = \frac{3}{2}u$, $z = \frac{1}{2}x^2$.

In case (b) take $x = 1 - \frac{1}{2}\log(3u-2)$, $z = \frac{1}{2}(x-2)^2$.

Select an exponential variate y (mean 1) and accept x if $y > z$.

2.6 Bivariate Sampling

We must now turn to the problems of sampling from bivariate distributions. This is considerably more difficult and only an outline of the procedures that can be used are given.

The 'Top Hat' Technique is of course available, in which the two variates in question are written on each disc. The mechanical extension to a two-dimensional tabular array can also be used. The difficulty here

is that the number of classes is likely to be large for any finely divided distribution and the number of discs or cells required becomes enormous.

The Conditional Probability Technique. This utilises the fact that we may write

$$p(x, y) = p(x)p(y|x)$$

where $p(x)$ is the marginal distribution of x and $p(y|x)$ is the conditional distribution of y, given that the other variate has the value x. We use one of the preceding techniques to sample the value of x and then use this value to select which distribution the second variate y shall be selected from. This is really only of value when the conditional probability distribution is of the same form for all values of x and the sampling may be done from a single distribution and a scale factor and shift of origin applied. Otherwise a separate distribution for each x value requires approximation and the amount of storage required becomes prohibitive.

The special case of the normal distribution is the most important application. Suppose x_1, x_2 is a sample from a bivariate normal distribution with means μ_1, μ_2, standard deviations σ_1, σ_2 and correlation ρ. Then the conditional distribution of x_2 given x_1 is normal with mean $\mu_2+\rho x_1$ and standard deviation $\sqrt{1-\rho^2}\sigma_2$.

Thus the sample procedure consists of taking two random variates y_1 and y_2, say, from the normal distribution with mean zero and standard deviation 1, and forming

$$x_1 = \mu_1+\sigma_1 y_1$$

$$x_2 = \mu_2+\rho x_1+\sqrt{1-\rho^2}\sigma_2 y_2$$

2.7 Multivariate Sampling

The difficulties of multivariate sampling are similar to those of bivariate sampling with the difficulties increased many-fold. The principal difficulty in empirical studies will be to obtain an adequate quantity of data to represent the multivariate distribution adequately. The 'Top Hat' technique or its mechanical version can be used as before, but the number of discs or cells will rise astronomically.

In theory, the conditional probability technique is still applicable by writing the multivariate distribution

$$p(x_1, x_2, x_3, x_4, \ldots, x_n) = p(x_1)p(x_2|x_1)p(x_3|x_1x_2)\ldots p(x_n|x_1, x_2, \ldots, x_{n-1})$$

but it is very rare that it is possible to describe in adequately compact form the various conditional distributions involved.

In the case of a multivariate normal distribution, an alternative method is available. In this case, for zero means the distribution is completely described by its variance matrix, a matrix whose elements describe the variances and co-variances of the different variates. We may then proceed as follows:

If the variance matrix is Ω, this is known to be positive definite and there exists a non-singular triangular matrix Γ such that

$$\Omega^{-1} = \Gamma\Gamma'$$

where Γ' is the transpose of Γ

Hence $$x\Omega^{-1}x' = (x\Gamma)(\Gamma x)'$$

and thus the variables $y = \Gamma' x$ are independent normally distributed with unit variances.

Thus the required variate may be obtained from the relation

$$x = \Gamma'^{-1}y$$

This may be applied to the two-dimensional case:

$$\Omega = \begin{bmatrix} \sigma_1^2, & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2, & \sigma_2^2 \end{bmatrix} \qquad \Omega^{-1} = \frac{1}{(\sigma_1\sigma_2)^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2, & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2, & \sigma_1^2 \end{bmatrix}$$

Hence

$$\Gamma = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}} \begin{bmatrix} \sigma_2, & -\rho\sigma_1 \\ 0, & \sigma_1\sqrt{1-\rho^2} \end{bmatrix} \qquad \Gamma^{-1} = \begin{bmatrix} \sigma_1\sqrt{1-\rho^2}, & 0 \\ \rho\sigma_1, & \sigma_2 \end{bmatrix}$$

$$x_1 = \sigma_1(\sqrt{1-\rho^2}\,y_1 + \rho y_2)$$

$$x_2 = \sigma_2 y_2$$

which is substantially equivalent to the formula of the preceding section.

2.8 Sampling under Dependence

The foregoing sections have discussed the drawing of independent samples from various distributions. In some processes, however, the distribution from which a sample is drawn depends on the preceding samples. This can be regarded as a special case of drawing from a multi-variate distribution with the difficulty that, in general, the dimensionality of the distributions increases as the sample is gathered.

This complicated situation is one in which the practical difficulties of gathering sufficient information to describe this sequence of distribution functions often makes its simulation impossible. However, in certain theoretical studies, it is often assumed that certain simple forms of the dependence of one distribution on the preceding values hold. The most useful case concerned is the Markov process, where the sequence of values $x_0, x_1, x_2, x_3, \ldots$ are generated by the following rule:

$$x_{r+1} = ax_r + \varepsilon_{r+1} \qquad r = 1, 2, 3, \ldots \tag{A}$$

where the ε_r are independent sample values from some given distribution. In this case, and its multivariate extension where

$$x_{r+1} - \sum_{s=0}^{m} a_s x_{r-s} = \varepsilon_{r+1} \tag{B}$$

can be dealt with by successively sampling from the distribution ε_r and substituting in the equations (A) or (B).

If this model is thought to be adequate to describe the dependence of any empirical data, the coefficients a_s of equation (B) can be estimated by an extension of the methods of least squares. This technique is often used to describe the variation of the economic variables.

CHAPTER 3

MORE SAMPLING METHODS

The methods outlined in the last chapter have been of general application, and quite independent of the distribution in question. However, it is possible to use alternative methods for some theoretical distributions and these are discussed briefly below.

3.1 Normal Distribution

A good approximation to a normal random deviate can be obtained by the use of the central limit theorem. This states that the mean of n random variables each from the same distribution will, in the limit as n tends to infinity, have a normal distribution. The standard deviation will be proportional to $\frac{1}{\sqrt{n}}$. Thus if n uniform variates (random numbers treated as fractions) are added together, the resulting variate has an approximate normal distribution with mean $\frac{1}{2}n$ and standard deviation $\sqrt{n/12}$. In practice, reasonable approximations can be found with values of n as low as 10.

Since it is possible to find theoretically the convolution of two uniform variates, the number of items to be added may be halved by sampling from that convolution. This has a triangular distribution with a distribution function given by

$$P(x) = \begin{cases} 2x^2 & x \leqq \frac{1}{2} \\ 1-2x^2 & x \geqq \frac{1}{2} \end{cases}$$

and thus, if the random number is P, the required transformation used to pick a variable from the triangular distribution is

$$P \leqq \tfrac{1}{2} \qquad x = \sqrt{\tfrac{1}{2}P} \qquad P \geqq \tfrac{1}{2} \qquad x = \sqrt{\tfrac{1}{2}(1-P)}$$

The relative merit of these two alternative methods of generating a normal variate depends upon the difficulty of picking a random number and that of taking the square root. Thus, if this is being performed on an automatic digital computer, the simpler process picking more random numbers will usually be quicker.

Teichroew has extended this technique by considering polynomial functions in a random variable θ

$$y = \sum_{r=1}^{n} a_r \theta^r$$

$$\theta = \sum_{i=1}^{m} \xi_i$$

ξ_i being uniformly distributed. He considers the number of terms to be used in forming θ and the optimum values of the coefficients a_r and has developed some powerful formulae for the efficient sampling of the normal distribution.

For example, if the normal distribution is truncated with a tail probability of 10^{-5} an adequate fit is given by $m = 8$, $n = 13$ and $a_{2r} = 0$. The values of a_{2r+1} are given in the following table

i	a_i
1	4·1082897
3	0·1491695
5	0·0214077
7	0·0041338
9	0·0008779
11	0·0001639
13	0·0000043

An interesting approximation for the distribution is given by Kahn. This derives from the approximation

$$e^{-\frac{1}{2}x^2} \simeq \frac{2\,e^{-kx}}{(1+e^{-kx})^2} \quad , \quad x > 0 \quad , \quad k = \sqrt{\frac{8}{\pi}}$$

It is not clear how this remarkable approximation was discovered, but it does give an adequate representation.

The c.d.f. corresponding to this p.d.f. is

$$\int_0^x \frac{2k\,e^{-kx}}{(1+e^{-kx})^2}\,dx = 2\int_y^1 \frac{dy}{(1+y)^2} = \left[\frac{2}{1+y}\right]_1^y = \frac{2}{1+e^{-kx}} - 1$$

Thus the inverse c.d.f. is

$$x = \frac{1}{k}\log_e \frac{1+y}{1-y}$$

A random sign is attached to this variable where y is a uniform random variable in [0, 1].

Another approximation is given by Hastings. Put

$$\eta = \{-2 \log (1-y)\}^{\frac{1}{2}}$$

Then a standardised normal deviate is approximately

$$x = \eta - \frac{a_0 + a_1\eta}{b_0 + b_1\eta + b_2\eta^2}$$

where

$$a_0 = 2{\cdot}30753 \qquad b_0 = 1{\cdot}00000$$
$$a_1 = 0{\cdot}27061 \qquad b_1 = 0{\cdot}99229$$
$$b_2 = 0{\cdot}04481$$

This applies if $y \geqq 0{\cdot}5$. If $y < 0{\cdot}5$, replace y by $1-y$ and change the sign of x.

A powerful method recently developed by Box and Muller generates pairs of normal deviates.

The joint distribution of two independent standardised normal deviates x_1, x_2 is

$$p(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$$

Consider the transformation

$$x_1 = r \cos \theta$$
$$x_2 = r \sin \theta$$

The general distribution of r, θ is

$$p(r, \theta)\, dr\, d\theta = \frac{1}{2\pi} e^{-r^2} r\, dr\, d\theta$$
$$= e^{-r^2} d(\tfrac{1}{2}r^2) \times \frac{1}{2\pi} d\theta$$

Thus r^2 and θ are independently distributed: θ has a uniform distribution in $(0, 2\pi)$ and r^2 has an exponential distribution.

Given two uniform random variables U_1, U_2

$$R = \sqrt{-\log_e U_1}$$
$$\theta = 2\pi U_2$$

are distributed as r and θ and hence

$$X = \sqrt{\log_e \frac{1}{U_1}} \cos 2\pi U_2$$

$$\sqrt{\log_e \frac{1}{U_1}} \sin 2\pi U_2$$

are *exact* independent normal deviates.

The value of this method depends on the facilities and speed of calculating logarithms and the trigonometric functions.

The usual trigonometric function routines usually work faster for small angles and a single deviate can be obtained by assigning a random sign to

$$X' = \sqrt{\log_e \frac{1}{U_1}} \sin \frac{\pi}{2} U_2$$

3.2 The χ^2 or Gamma Distribution

The samples from this distribution may be obtained by using the fact that χ^2 on n degrees of freedom is the convolution of n variables distributed as χ^2 on one degree of freedom, which is in turn merely the square of a normal random deviate. Thus, to pick a sample value from χ^2 on r degrees of freedom, r random normal deviates are selected, squared and summed.

It is often necessary to sample from χ^2 on n degrees of freedom where n is not fixed on successive occasions, but is determined either by a random process or by other considerations. In this case, the method outlined above is still available, but can be shortened if the lower bound to the value of n to appear is known. Suppose this is n_0 then we may take a sample from χ^2 on n_0 degrees of freedom using the transformation method and pick $m = n - n_0$ normal deviates $x, x_2, \ldots, x_m$ and form

$$\chi_n^2 = \chi_{n_0}{}^2 + \sum_{i=1}^{m} x_i^2$$

Naturally n_0 must be a lower bound, because if otherwise, m will be negative. It is not possible to obtain the required distribution by subtracting the squares of random normal deviates since

$$\chi_n^2 - \chi_m^2 \neq \chi_{n-m}^2 \qquad m < n$$

This is easily seen, since the only quadratic forms of normal variates that have χ^2 distribution have matrices with idempotent eigenvalues.

A quicker method hinges on the fact that an exponential distribution is χ^2 on 2 degrees of freedom.

Thus for even degrees of freedom ($n = 2p$) we take

$$\chi_n^2 = \sum_{i=1}^{p} U_i$$

where the U_i are exponentially distributed.

For odd degrees of freedom $n = 2p + 1$ we take

$$\chi_n^2 = \sum_{i=1}^{p} U_i + Z^2$$

where Z is normally distributed.

3.3 Exponential Distribution

It has already been mentioned that this distribution may be found merely by taking a scaled value of the logarithm of a random number.

Von Neumann has suggested a rejection technique for selecting from an exponential distribution.

Choose two random numbers R_0, R_1.

If $R_1 \leqq R_0$ score 1 and choose a new random number R_2.
If $R_2 \leqq R_1$ score 1 „ „ „ „ „ „ R_3.

Continue in this way until the test $R_i \leqq R_{i-1}$ fails.

The joint distribution of the score I and the variable R_0 is obtained as follows.

If exactly i tests are made $R_0 \geqq R_1 \geqq R_2 \ldots \geqq R_i < R_{i+1}$ and the probability of this can be evaluated as

$$\frac{R_0^i}{i!} - \frac{R_0^{i+1}}{(i+1)!}$$

The probability of $I = i$ is $\dfrac{1}{(i+1)!} - \dfrac{1}{(i+2)!}$

For given R_0 the probability of i even is

$$\left(1 - \frac{R_0}{1}\right) + \left(\frac{R_0^2}{2!} - \frac{R_0^3}{3!}\right) = \Sigma(-)^r \frac{R_0^r}{} = e^{-R_0}$$

The conditional probability of R_0 for i even is

$$\frac{e^{-R_0}}{\int_0^1 e^{-R}\,dR} = (1-e^{-1})^{-1}\,e^{-R_0}$$

Thus if the sequence of tests is repeated until failure occurs on an even i the initial random number has been selected from the truncated distribution $(1-e^{-1})^{-1}\,e^{-x}$.

The number of sequences n tried before success has a distribution $(1-e^{-1})\,e^{-n}$.

The variables n and R are independently distributed so the p.d.f. of $x = n+R$ is given by e^{-x}.

3.4 The Poisson Distribution

If the intervals between events are exponentially distributed, it is well known that the number of events in a fixed period of time has the Poisson distribution, and this fact may be used to generate such a distribution. The Poisson distribution, of course, is a frequently occurring example of a discrete distribution and is given by

$$P(x = n) = \frac{\lambda^n}{n!}\,e^{-\lambda} \qquad n = 0, 1, 2, \ldots$$

We construct a sequence y_1, y_2, y_3 of variables with an exponential distribution and mean $\frac{1}{\lambda}$ and form their cumulative sum. Let n be defined by

$$\sum_{i=1}^{n} y_i \leqq \lambda < \sum_{i=1}^{n+1} y_i$$

Then n will have a Poisson distribution with mean λ. Of course, if λ is large, the normal approximation can be used.

An alternative technique based on rejection is also available.

Consider the distribution of the product R_0R_1 of two uniform random variables.

For fixed R_0 $$R_1 = \frac{x}{R_0}$$

$$\therefore \quad p(x) = -\log x$$

By induction the product $\prod_{i=0}^{n} R_i$ has p.d.f.

$$f_n(x) = (-\log x)^n/n! \qquad 0 \leqq x \leqq 1$$
$$= 0 \qquad \text{otherwise}$$

and c.d.f. $$F_n(x) = x \sum_{i=0}^{n} \frac{(-\log x)^n}{n!}$$

Take the product U_n of n uniform random variables and test

$$U_n > e^{-m}$$

The probability of this is $1-F_n(e^{-m}) = 1-e^{-m} \sum_{j=0}^{n} \frac{m^j}{j!}$

The distribution of the first n for which $U_n > e^{-m}$ is

$$F_n(e^{-m}) - F_{n-1}(e^{-m}) = \frac{m^n}{n!} e^{-m}$$

i.e. n has a Poisson distribution of index m.

3.5 Other Rejection Techniques

Kahn lists a number of other distributions which can be sampled by rejection techniques. A few examples follow:

(1) To sample from the p.d.f. given by

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}} \qquad 0 \leqq x \leqq 1$$
$$0 \qquad \text{otherwise}$$

This is the distribution of $\cos \pi R$ where R is a uniform variate in [0, 1].

(i) Select two uniform variates R_0, R_1

(ii) Reject unless $R_0^2 + R_1^2 \leqq 1$

(iii) Now (R_0, R_1) is a random point in the first quadrant of a circle and

$\arg \eta = \tan^{-1} \frac{y}{x}$ is uniform in $\left[0, \frac{\pi}{2}\right]$

The required variate is

$$\cos 2\eta = \cos^2 \eta - \sin^2 \eta = \frac{R_0^2 - R_1^2}{R_0^2 + R_1^2}$$

In practice, if a pair is rejected, the last value can be retained for the next trial, reducing the number of random numbers needed.

The efficiency of this process is obviously $\frac{\pi}{4}$.

(2) The following rejection procedure can be used to generate a number of different variates.

(i) Set $i = 0$, select a uniform random variable R_0.
(ii) Select a random variable R_{i+1}, increase i by 1.
(iii) Test if $R_i \geqq R_{i-1}$; if so repeat stage (ii);

if not set

$$X = R_{i-1}$$
$$Y = R_i$$
$$Z = R_0$$
$$I = i$$

The joint p.d.f.'s for (X, I), (Y, I), (Z, I) can be obtained as

$$f(x, i) = \frac{x^i}{(i-1)!} \quad , \quad 0 \leqq x \leqq 1 \quad , \quad i \geqq 0$$

$$g(y, i) = \frac{1-y^i}{i!} \quad , \quad 0 \leqq y \leqq 1 \qquad i \geqq 0$$

$$h(z, i) = \frac{(1-z)^{i-1}}{(i-1)!} - \frac{(1-z)^i}{i!}$$

The marginal p.d.f. for i is obtained as

$$Pr(I = i) = \frac{1}{i!} - \frac{1}{(i+1)!}$$

But the marginal distributions of X and Y are of more interest

$$f(x) = \sum_0^\infty \frac{x^i}{(i-1)!} = xe^x \quad , \quad 0 \leqq x \leqq 1$$

$$g(y) = e - ye^y \qquad , \qquad 0 \leqq y \leqq 1$$

That for Z is, of course, merely the original uniform distribution.
If samples which stop on even values of i are rejected, the marginal distributions become

$$f_0(x) = \frac{1}{K}\sum_0^\infty \frac{x^{2i+1}}{(2_i)!} = \frac{x\cosh x}{K} \quad \text{where} \quad K = 1-e^{-1} \quad , \quad 0 \leqq x \leqq 1$$

$$g_0(y) = \frac{(\sinh 1 - \sinh y)}{K} \qquad 0 \leqq y \leqq 1$$

$$h_0(z) = \frac{e^z}{e-1} \quad , \qquad 0 \leqq z \leqq 1$$

If samples which stop on odd values of i are rejected, the marginal distributions are

$$f_e(x) = ex\sinh x \quad , \qquad 0 \leqq x \leqq 1$$

$$g_e(x) = e(\cosh 1 - \cosh y) \quad , \quad 0 \leqq y \leqq 1$$

$$h_e(x) = e - e^z \quad , \qquad 0 \leqq z \leqq 1$$

Replacing the test $R_i \geqq R_{i-1}$ by $R_{i+1} \geqq R_i$ replaces the variables x, y, z by their complements. This enables a further set of distributions to be generated.

A further set can be generated by replacing the test by $R_i < R_0$ and yet another by $(i+1)R_i < iR_0$. For further details Kahn should be consulted.

The whole process is an extension of the Von Neumann technique for generating an exponentially distributed variable.

3.6 The Binomial Distribution

This distribution is the most frequently required discrete distribution, and there are several special methods of generating variables from this. For the special case of $P = \frac{1}{2}$ the digits of a uniform random variable are variables which take the values 0 and 1 with equal probabilities $\frac{1}{2}$. Thus the binomial variable of index n with $P = \frac{1}{2}$ can be formed by summing n digits of a uniform random variable.

For the case of $P \neq \frac{1}{2}$ this method can be extended and is best explained by an example.

Suppose we require a sample from a binomial distribution with $P = \frac{1}{4}$, then take two random variables x_1, x_2 and express them in the binary scale, e.g.

$$x_1 = \cdot 11010111001011001$$

$$x_2 = \cdot 10101101100110101$$

If we now form a new number by a process of digitwise logical multiplication of these two numbers we obtain the number

$$\cdot 10000101000010001$$

The occurrence of a 1 will only arise with probability of $\frac{1}{4}$. Hence the sum of n digits from the collation of two uniform random variables gives the required random variable from a binomial distribution of index n and $P = \frac{1}{4}$. If a value of $P = \frac{3}{4}$ is required, then we complement the number given by the collation, and sum n digits of this number.

By a suitable sequence of collations and complementations we can obtain a sequence of digits in which the probability of a 1 may have any required value, expressed as a binary fraction. Thus to obtain a sequence with $p = \frac{3}{16}$ we take four random numbers, collate the first pair and complement giving a sequence with $p = \frac{3}{4}$, collate the second pair giving a sequence with $p = \frac{1}{4}$ and collate the two results.

The advantage of this technique is that several random variates with the required distribution can be obtained from a single random number or group of random numbers.

A similar device has been described by Davis as follows. Treat the random number x as a fraction.

If $x < p$, score 1 and form $x_2 = x_1$.

If $x \geqq p$, score 0 and form $x_2 = x_1 - p$.

The probability of scoring 1 is p and x_2 is uniform in range $(0, p)$ or $(0, q)$. We could rescale x_2 to be uniform in 0, 1 and repeat, but the division can be avoided by rescaling the probability p to p^2 or pq $(q = 1-p)$.

Thus a repetitive process can be defined to give a sum of scores.

$$x_1 = x$$

$$p_1 = p$$

If $x_i < p_i$, score 1 and form $x_{i+1} = x_i$

$$p_{i+1} = pp_i$$

If $x_i \geqq p_i$, score 0 and form $x_{i+1} = x_i - p_i$

$$p_{i+1} = qp_i$$

This process may be continued until the successive multiplications have reduced the size of x to too few significant digits for the comparison to be accurate.

3.7 Selecting a Point at Random in an n-dimensional Volume

If a geometric problem is being studied by simulation techniques the need to select a point at random in n-space is required. If the space is an n-dimensional cube of unit size the problem is trivial since the point $(x_1, x_2, \ldots, x_n)$ has the property that $x_1, x_2, \ldots, x$ are independent uniform variates.

To select a random from any other bounded space V the rejection process can be used. Surmount the space by a larger cube, sample from this enclosing space and test if point is in V, if not reject it. The first accepted point is a randomly chosen point in V.

In two and three dimensions, this process is quite satisfactory, but in higher dimensions the rejection rate becomes very high. For example, if V is an n-dimensional sphere of unit radius the probability of acceptance of a point randomly chosen in the hyper-cube $-1 \leqq x_i \leqq 1$ $(i = 1, 2, \ldots, n)$ is

$$P_n = \frac{\pi^{n/2}}{n2^{n-1}\Gamma(n/2)} \sim k\left(\frac{\pi e}{4n}\right)^{n/2}$$

and decreases alarmingly as n reaches values of 10–20. The rejection process then becomes impracticable.

An alternative procedure consists of choosing a radius x $(\leqq 1)$ and selecting a point on the surface of the hyper-sphere of radius x at random. Since the surface area of a hyper-sphere of radius x is

$$\frac{2\pi^{n/2}x^{n-1}}{\Gamma(n/2)}$$

x must be sampled from a distribution with p.d.f.

$$f(x) = nx^{n-1} \quad , \quad 0 \leqq x \leqq 1$$
$$= 0 \quad , \quad \text{otherwise}$$

The c.d.f. is

$$F(x) = x^n \quad , \quad 0 \leqq x \leqq 1$$
$$= 0 \quad , \quad \text{otherwise}$$

and the usual inverse process gives

$$x = \sqrt[n]{y}$$

where y is a uniform variate.

To pick a point at random on the surface of a hyper-sphere requires the selection of a direction at random. Pick $x_1, x_2, \ldots, x_n$ from a normal distribution.

Their joint distribution has p.d.f.

$$\left(\frac{1}{\frac{1}{2}\pi}\right)^{n/2} e^{-\frac{1}{2}(x_1^2+x_2^2 \ldots x_n^2)}\, dx_1\, dx \ldots$$

and is radially symmetric.

Thus $(\xi x_1, \xi x_2, \xi x_3, \ldots, \xi x_n)$ is a point on the n-sphere if

$$\xi = \frac{x}{(x_1^2+x_2^2+x_3^2 \ldots x_n^2)^{\frac{1}{2}}}$$

An alternative procedure determines a point on an n-sphere by a recursive procedure, selecting a point on a 1-sphere ($x_1 = 1$ or $x_1 = -1$) and determining from this a random point on a 2-sphere and so on until the required dimensionability is reached.

Given a point $(x_1, x_2, \ldots, x_m)$ on an m-sphere, the point on the $(m+1)$ sphere is

$$(rx_1, rx_2, \ldots, rx_m, s\sqrt{1-r^2})$$

where s is a random sign and r has the p.d.f.

$$k\frac{r^{m-1}}{\sqrt{1-r^2}} \quad , \quad 0 \leqq r \leqq 1 \quad , \quad k = \left(\int_0^1 \frac{r^{m-1}}{\sqrt{1-r^2}}\, dr\right)^{-1}$$

$$0 \quad , \quad \text{otherwise.}$$

The first method can be adapted quite easily to sample uniformly within an n-dimensional ellipsoid.

CHAPTER 4

RANDOM NUMBER TABLES

4.1 Introduction and History

In Chapter 2 we reduced the problem of sampling from a distribution to one of sampling from a uniform distribution (or if one is working arithmetically to picking an integer at random from a consecutive range). When sampling experiments are being performed manually, the main technique for ensuring that the numbers used are random consists of generating a set of numbers by some random process, tabulating them and using the tables on the various successive jobs of sampling undertaken.

The modern history of sampling experiments in mathematical statistics started with the work of W. E. Gossett writing under the pseudonym of 'Student' from Guinness's Brewery in Belfast. He discovered the celebrated t-distribution named after him, and at the time he published the theoretical derivation of the distribution of this statistic, he backed his findings by sampling experiments.

This technique of discovering or verifying the sampling distribution of interesting statistics was taken up in a vigorous fashion by the statistics school at the University College under Karl Pearson, and for several years a steady stream of papers was published giving the results of such experiments. In all these, the sampling was performed by the traditional 'Top-Hat' method using beads or discs in an urn or bag. The idea of using tables of random numbers was introduced by Tippett, when working in their laboratory on the distribution of the range of a sample from a normal distribution. To perform this work on the range, Tippett constructed a table of 10,400 random digits which he extracted as the terminal digits of entries in some census tables.

This gave a new impetus to sampling experiments, which could now be performed much more quickly and with greater reliability than before. It soon became clear that the 10,400 numbers in Tippett's tables were not sufficient, and M. G. Kendall and B. Babington-Smith, working in co-operation with University College, decided to form a far more extensive table. They produced a table of 100,000 decimal digits, using a mechanical randomiser to generate them. This will be described

in the next chapter. At about the same time, Fisher and Yates, working from Rothamsted, produced a smaller table of 15,000 digits which were in fact the 15th to the 19th digits of the logarithms given in A. J. Thompson's mammoth *Logarithmica Britannica*.

This set of tables sufficed for sampling experiments for many years, and it was not until after the Second World War that any significant addition to the stock of random number tables was made. The Rand Corporation in America produced 1,000,000 random digits, using an electronic randomiser with its own printing device. This machine will also be described in the next chapter.

4.2 Tests of Random Number Tables

The authors of all these tables held some doubts of the validity of their random sources and consequently determined to make some tests that these tables actually constituted a sample of random numbers. Kendall and Babington-Smith, in the paper in which they described their tables, also give an account of the tests which they applied to their tables and these have become, through the years, the standard tests applied for randomness. Kendall and Babington-Smith emphasised that the number of tests for randomness that can be applied to a sequence of digits is quite unlimited except by human ingenuity, and implied that it will always be possible to invent some test that a given table will fail, and in their paper they give an interesting philosophical discussion of the implications of such tests. Broadly speaking, their conclusions are that the tests of randomness applied to a sequence will depend upon the use to which this sequence is to be put. If it can be established that a particular type of patterning in the table is likely to give serious error to the results of some sampling experiment, then the table must be tested so that this particular type of patterning does not exist. This tailor-made approach, of course, is not very realistic in practice, and the custom has arisen of accepting the tests suggested by Kendall and Babington-Smith for nearly all purposes. These tests are now described.

4.3 Frequency Test

The first obvious requirement of a set of decimal digits is that each decimal digit shall occur with approximately equal frequency, and this test merely consists of recording the frequency of occurrence in the table of each decimal digit and comparing with its expected frequency (one-tenth the sampled size) by using a χ^2 test.

4.4 The Serial Test

The frequency test merely tests the probability of occurrence of each digit in a given position, but it does not exclude the possibility of serial correlation between the digits in successive positions. Thus the sequence 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, ... will satisfy the frequency test, but clearly it is not a random sequence. As a safeguard against serial correlations between successive digits, Kendall and Babington-Smith applied a test in which the frequencies of every pair of digits were recorded and tested against expected value. In this test, each digit is used twice, first as a leading digit, second as a trailing digit, and a comparison with the actual frequencies with the expected are once again made with a χ^2 test.

4.5 The Poker Test

The serial test examines first order correlations only and for higher order serial correlations some more comprehensive test is required. This could be achieved by recording the frequencies of triads of digits compared with expectation and then with quadruples and quintuples. However, for large groups of this kind, the number of values contributing to the contingency table will be severely restricted, and the numbers in each class may become rather small to the consequent detriment of the effectiveness or the validity of the χ^2 test. To overcome this, Kendall and Babington-Smith proposed to amalgamate certain of the five digit groups and make a test of these composite groups. The groups chosen were selected according to the principles of poker. The groups were those in which

(i) all digits were distinct;
(ii) a pair of like digits occurred;
(iii) two pairs of like digits occurred;
(iv) three like digits occurred;
(v) four like digits occurred; and
(vi) all digits were alike.

It is a simple exercise in combinatorial arithmetic to work out the expected frequencies in these classes and apply the χ^2 test for goodness of fit of the actual distribution to the theoretical one.

Because of this similarity between the composite groups and those in the card game of poker, it is known as the poker test. In this test, the whole sample n is divided into $n/5$ samples of quintuplets so that the problems of dependence between the different groups is eliminated.

4.6 The Gap Test

The final test as suggested by Kendall and Babington-Smith was that the number of digits between successive zeros should be counted and the frequency of the various 'gaps' be compared with the expected values given by the ordinary geometric series.

Kendall and Babington-Smith applied the χ^2 method of testing goodness of fit in all these cases, but it has been pointed out by Good that this is not valid for the serial test, as the use of each digit twice introduces a bias. He has shown that the expected value of the χ^2 statistic calculated in this test will be 99, whereas, bearing in mind that there are 10 linear restraints, it should have an expected value of 90. This difficulty can be circumvented at the expense of the stringency of the test by taking the digits in n pairs, and therefore using each digit once only. Although the practical implications of Good's analysis are not very serious for the tests of decimal digits, they become very much more serious in the case of binary digits where there are only four classes in the serial test (00, 01, 10 and 11) and the χ^2 calculated would have an expected value of 3 as opposed to the presumed value of 2. However, if tables of binary values were produced, or if the sequences of binary values produced by digital computers were to be tested, the volume of digits available is likely to be much higher, and then a test on the relative frequencies of all possible quintuplets could be made, dividing the sequence up into independent 5-bit parts.

Other tests have been described from time to time in literature, and a short account of these is now given.

4.7 The Run Test

Kermack and McKendrick have suggested a test of random digits using a run property. They defined a run as a sequence of digits either in ascending or descending order, and the number of digits in a run is termed the length of the run. A descending run must, by definition, be followed by an ascending run and vice versa, and the length of the run down and up (or up and down) is termed a gap. The derivation of the distribution of the run and gap length is rather difficult, and in their original paper, they restrict themselves to the problem of evaluating this distribution for a randomly arranged set of unequal numbers. In the case of a table of random digits, of course, each digit may appear many times, and the possibility of ties in the sequence considerably complicates the analysis. For this reason, presumably, this test has

never received much attention, and is not today used for the testing of random sequences.

4.8 Yule's Test

Yule, in using Tippett's tables, noticed a certain type of patchiness in the digits which he suspected affected his results, and in consequence he applied a test of his own invention. This consisted of examining the actual distribution of the sums of five digits with its expected distribution. He chose the five elements to be added by reading down the columns of the table, and showed that certain batches of Tippett's tables did not satisfy this test of randomness. The theoretical distribution of the sums of five equally likely digits presents some little theoretical difficulty, which Yule overcame by a practical arithmetic process. He derived it in five stages: first finding the distribution of the sum of two digits, then the distribution of the sum of two digits, and one digit, and so on.

4.9 The D^2 Test

Random numbers are now often used for the evaluation of multi-dimensional integrals. This is regarded as a generalisation of the problem of evaluating an area. This uses random numbers to select a point at random in a unit square surrounding the area. When the numbers are to be used in this way, it would be sensible to test that the points so derived are uniformly distributed over the square. A test of this feature of the sequence has been devised by Gruenberger and Mark that consists of using sequences of four random numbers as the two pairs of co-ordinates of two points in a square and computing the distance between them. The distribution of this distance has a theoretical distribution that can be easily calculated from the formulae

$$P(d) = \pi d^2 - \frac{8d^3}{3} + \frac{d^4}{2} \quad , \quad 0 \leqq d \leqq 1$$

$$= \tfrac{1}{3} + (\pi - 2)d^2 + 4(d^2-1)^{\frac{1}{2}} + \tfrac{8}{3}(d^2-1)^{\frac{3}{2}} - \tfrac{1}{2}d^4 - 4d^2 \sec^{-1}$$

$$1 < d \leqq \sqrt{2}$$

and, as usual, the actual can be compared with the expected by means of a χ^2 test.

4.10 Comments Concerning Tests

These tests are usually made on a collection of random digits in fairly small batches. It is possible then by combining the χ^2-value from

each of the batches to make an overall test of the whole table. If a table is large enough, then we must expect small parts of the table to exhibit highly regular patterns as any particular regular pattern must occur with exactly the same frequency as any other configuration. Thus we must not be surprised if a table of $10^{10^{10}}$ digits contains at least one example of any one sequence of 10^{10} digits. In particular, a whole series of zeros should occur. Such regular patterns will give high χ^2 values in some or all of the tests. In fact, if the table is broken into small batches and for any given test the χ^2 value is calculated for each batch, then the set of χ^2 values obtained should actually be a sample from the appropriate χ^2 distribution. Some of the later workers have tested their χ^2 values from separate batches in this way. Regarded in the large, any table of random numbers which does not include its due proportion of high χ^2 valued batches cannot be regarded as a random table. On the other hand, for practical use, that particular batch of the table would be quite useless in some small limited experiment.

Thus the compilers of tables have usually described the tests that have been made on them and the results of those tests, and indicate which of the various batches in their table fail to meet each of the several tests applied.

4.11 Tables of Other Random Variates

Although we have described in some detail in the preceding chapter how to convert from random numbers to any other statistical distribution, certain distributions are of such importance that tables of samples from such distributions can save an enormous amount of work in sampling experiments. Of paramount importance in statistical theory is the normal distribution, and two tables of random normal deviates have been published. These were both found by transforming tables of random numbers by means of the inverse cumulative formula given in Chapter 2. Mahalonobis transformed Tippett's tables, and Herman Wold of Uppsala transformed the Kendall and Babington-Smith tables into 25,000 random normal deviates. Both these tables, when constructed, were subjected to further tests for their randomness. The test used consisted of forming the distributions of the sum of n of these random normal deviates, and comparing this distribution with the expected normal distribution.

4.12 Conclusion

The tables of random numbers described in this chapter were historically of the utmost importance in the advancing of the technique

of sampling experiments, and are still an invaluable aid to workers who wish to make small-scale experiments. However, the advent of electronic computers has enabled the scale of operations to be considerably enlarged, and these tables can now no longer be thought of as a good source of randomness for the enlarged sampling experiments being conducted. The next two chapters consider two alternative sources of randomness for large-scale experiments. The first of these is the physical generation of random digits by some apparatus, and the second is the production of sequences of digits that, although produced deterministically, will satisfy the various tests for randomness that have hitherto been applied to tables.

CHAPTER 5

RANDOM NUMBER GENERATORS

5.1 Introduction

We have seen that the method of simulation (or sampling experiments) requires a source of random digits which can be used singly or in collections as approximations to uniform random variates. Chapters 2 and 3 were concerned with the use of such a source of random numbers for the generation of samples from any statistical distribution. This chapter will consider methods of deriving such random digits by physical means.

There is a long history of building machines to generate random digits. This arose, in the first instance, from the need to study certain stochastic phenomena by building physical models of the systems. The present-day physical generators of random digits have derived from these attempts to build physical random simulators.

The first and most important machines constructed for physical simulation were used by telephone traffic engineers to study the impact of random fluctuations of telephone demand on telephone exchanges, and the efficiency of various methods of routing calls through the available equipment to minimise the delays to those calls. Similar machines have been made to study the analogous problem of random effects on road traffic congestion.

A wide variety of natural phenomena have been used to produce randomness and can be divided into the electronic and non-electronic devices. Amongst the non-electronic or mechanical devices, the most popular has been the application of the spinning disc used in many fair-ground games. A disc, divided into a set of equal numbered sectors, is spun and allowed to come to rest. The number of the sector selected by a fixed pointer is taken as the random number selected.

Most electronic devices make use of the small variations in the thermionic emission from a valve due to the thermal agitation of the electrons in the filament. Another popular source of noise is the emission from a radio-active source.

It has been proposed to use detection apparatus to pick up atmospheric radiation and convert this to a widely fluctuating voltage.

Another possibility is to utilise the Brownian movement and detect the incidence of a Brownian particle in a small field by means of photo-electric devices. A machine based on this principle is known to have been constructed.

Controversy has raged about the validity of each of these methods and philosophers have argued whether such natural phenomena can be regarded as random. For practical purposes these arguments are irrelevant, and it seems that the right requirement of a phenomenon to be suitable for a random noise source is that there is no scientific explanation of the phenomena that is used and the effect that is observed is generated by a large number of atomic events. Moreover, these atomic events, in spite of their large numbers, must not display any of the statistical regularity that is observed as, say, in the gas laws.

The crucial test that a process is random is that predictions concerning the future behaviour of the system cannot be improved from a knowledge of its past behaviour. Ideally, we should require proof that no rule for prediction is possible, but since there is no limit to the rules that can be tested, and ultimately the proof must depend upon experiment, this will not be possible. In practice, then, we are forced to accept any phenomena whose exact mechanism is unknown to us and whose behaviour is not predictable by any obvious deterministic laws.

It has already been mentioned that the second major table of random numbers was produced by Kendall and Babington-Smith, using a randomising device. This was a mechanical disc mechanism. A disc was spun by an electric motor and was stopped at an arbitrary time by the observer and the digit against the pointer noted for inclusion in the table.

The arbitrary time was selected by blindly drawing an electrode across a graphite maze drawn upon a piece of paper until the electrode made electrical contact with the maze.

The principle of this machine is that there will be many revolutions of the disc between observations and the only relevant factor in determining the digit recorded is the magnitude of the fractional part of the last revolution made. It is maintained that this fractional part is not under the control of the operator and so is equally likely to take any value and hence each of the ten sectors are equally likely to be selected. Kendall and Babington-Smith tested the tables produced by this mechanism by the several tests described in Chapter 4, and indeed (as was mentioned there) they were largely responsible for the invention of many of those tests.

Other mechanical randomisers have been built, but there is very little in the literature concerning them. The only exception to this is the stochastic analogue machine devised by Beer. This employs the principle of buffeting a ball in a random manner before allowing it to fall into one of a large number of pockets. This machine will be described in more detail later.

5.2 The Electronic Machines

These machines are based almost universally on converting the source of random noise into a train of pulses. This train of pulses is used to drive a cyclic counter. The drive is interrupted at fixed intervals of time and its state examined. The state between successive inspections of the counter depends on the number of pulses occurring in that interval of time; and more particularly on the remainder on dividing that number by the radix of the counter. Thus this mechanism is similar to that used with the spinning disc except that the roles of randomness and determinacy are interchanged. In the disc the inspection occurs at random intervals, and the events (change of state) are at constant intervals determined by the constant speed of the motor; in the electronic devices the inspection is at constant intervals and the events occur at random intervals.

Cursory consideration of these two systems would lead one to believe that they would be equally efficient, but it has been shown by Thompson, one of the workers on the G.P.O. machine ERNIE, that this is not the case. He has given a mathematical analysis of these two situations, which shows that the process of uniform inspection of a randomly actuated counter gives rise to a more accurate random distribution of the positions of the counter than the other system.

For those readers who would prefer a heuristic argument rather than a more mathematical one, the following may be useful.

Although the number of pulses between inspections has a discrete distribution, if the average number of pulses per sampling period is high enough, it can be regarded as a continuous distribution. Divide the distribution into vertical strips whose width corresponds to the radix of the counter. Then the distribution of relative movement of the counter is obtained by superimposing these strips on to each other. Each strip has a curved top, but for reasonably symmetric distributions, each strip may be mated with a symmetrically disposed strip and so the top of the summed strips is level, i.e. the distribution is uniform.

The wider the original distribution, the flatter the tops of the larger

strips are and the less levelling out is required by the summing process. Thus high variability is a desirable feature of the distribution, and this is the reason for favouring a Poisson source of pulses.

If the successive movements are uniformly distributed, then the successive counter positions are uniformly distributed. To ensure that the positions are statistically independent, the number of pulses in successive intervals must be independently distributed.

To ensure this, we require that the probability of generating a pulse at any instant does not depend on the history of the previous pulses generated. The random noise source from which pulses are generated can be analysed by means of auto-correlation functions, and it is fairly easy to see that for no dependence of one pulse train on another we require the auto-correlation coefficient for lags greater than or equal to the sampling interval to be zero. This requirement is met if the sampling interval is chosen large enough to exceed the maximum lag for which the auto-correlation of the originating noise source is significant.

To obtain a pulse train from the noise source, this is fed into a high-gain amplifier whose output is limited both ends, thereby producing an irregular square wave form. This square wave form is differentiated so that each swing of the amplified noise generates a pulse.

In practice, the differentiator in such a pulse-generating circuit will have a dead-time associated with it so that after generating a pulse it cannot produce another one for a certain interval of time. Extensive mathematical analyses have been made of the performance of systems involving dead-times in connection with counting the radio-active particles and the use of Geiger counters as measures of the strength of radio-activity. There is no need to go into the details of these analyses, since, for the purposes of constructing a random number generator, we rely upon the production of a long train of irregularly spaced pulses and counting them. The exact distribution of the intervals between them is now no longer of significance.

We now describe, in more detail, some of the actual machines that have been constructed.

5.3 ERNIE

This machine has obtained popular notice for its use in connection with the national lottery system recently introduced in this country. The randomising technique used is precisely the one described above and uses neon discharge tubes as the sources of noise. Each time the noise wave forms pass up above a certain fixed level, the pulse generators

emit a short pulse of two microseconds duration, followed by a dead-time of 30 microseconds. The cyclic counters used have 10 or 24 positions, the latter being used to select a random letter and the former for random decimal digits. The counters are inspected at intervals of about one-sixth of a second to produce the random digits.

The output from these generators should all be independent of each other, which implies that the noise wave forms from the neons are independent, and it is thus important to ensure that there are no common effects caused by fluctuating anode voltages.

For a machine used for the purposes of ERNIE, it is most important that no fault should develop during the selection of bond numbers, and for this purpose a redundancy technique has been used in the generation of the random digits. Nine random numbers are required, but ten cyclic counters are provided. These are connected in pairs, as is shown in Figure 6.

The final outputs are taken as the differences of each pair of generators. This has the effect of increasing the variability of the number of pulses counted, and hence leads to an improvement in the approach to equi-probability.

Its main purpose, however, is to deal with the possibility of the breakdown of one of the original generators. If, say, $U2$ breaks down, then $V1 = U1 - K$, where K is the constant value from the broken generator. Thus, if $U1$ is uniformly distributed, so is $V1$. Moreover, if K is not fixed, but has any distribution whatsoever, it is still true that $V1$ has a uniform distribution. Thus the nine outputs of the machine are safeguarded against the breakdown of any one of the ten generators.

The arrangement preserves the independence of the separate outputs since if the ten cyclic counter outputs are $r_1, r_2, \ldots, r_{10}$ and the final outputs are $s_1, s_2, \ldots, s_9$, we have

$$\begin{aligned} Pr(s_i = k_1, s_{i+1} = k_2) &= Pr(r_i - r_{i+1} = k_1, r_{i+1} - r_{i+2} = k_2) \\ &= Pr(r_i = k_1 + r_{i+1}, r_{i+2} = r_{i+1} - k_2) \\ &= Pr(r_i = k_1', r_{i+2} = k_2') = 10^{-2} \\ &= Pr(s_i = k_1) Pr(s_{i+1} = k_2) \end{aligned}$$

This principle could be extended to safeguard against more than one breakdown, and the interested reader is referred to the discussion on Thompson's paper for an analysis of the methods by which this can be achieved.

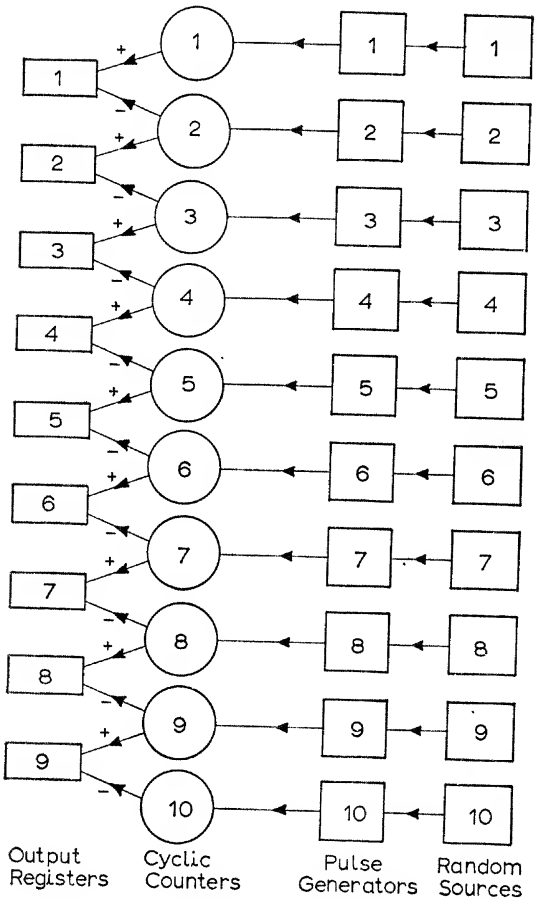


Figure 6

This machine is not typical of random number generators in that its output has a special application, and therefore a special form. It was developed from the traffic machines built by the G.P.O., which employed the same physical means of generation.

5.4 The Rand Corporation Machine

The Rand Corporation of America have recently published a table of one million random digits, which were produced on an electronic randomiser. This machine was of standard pattern and generated

decimal digits at the rate of one per second and printed the value obtained.

5.5 The Manchester Mark I Computer Randomising Unit

The first automatic computer built by Ferranti Limited (and designed by the Department of Electrical Engineering of the University of Manchester) incorporated a random number generator which used the standard principle. This machine worked in the binary scale and so had a binary counter (more usually known as a 'flip-flop') having only two states. This was continuously complemented by a train of pulses generated from a noise source, which, in this case, was the thermal agitation in the emission from a thermionic diode. The state of the flip-flop was inspected at times governed by the master clock of the machine, and the resulting digits collected into a register to form a random binary number. No details of this mechanism nor any descriptions of simulation experiments on the Manchester Mark I machine have, using this randomiser, ever been published to the author's knowledge. Further Ferranti machines have not incorporated such random number generators. The reader may draw his own conclusions.

5.6 The United Steel Companies Random Number Generator

This machine is a fast generator which has been built for experiments on cybernetic techniques of control of general systems. It uses the standard principles for the generation of random digits and, like the Manchester Mark I computer, works in the binary scale. The irregular square wave formed from the noise in a thermionic diode is fed directly to a binary counter, which is complemented by each rising edge of the wave form.

As in ERNIE, a comparison technique is used, but instead of using two random sources with identical characteristics, the same random number generator is used twice over. This is effected by supplying two binary counters, and a control mechanism which goes through the following cycle. First a gate 1 is opened, allowing the pulses to feed into counter 1, then gate 1 is shut and gate 2 opened, allowing the pulses to feed into counter 2, gate 2 is shut and an inspection gate 3, leading to an anti-equivalent circuit between states of counters 1 and 2, is opened. If these digits are unequal, the digit is recorded in a shifting register as that of the state of counter 1. If they are equal, no digit is recorded and the process is repeated. Figure 7 illustrates the process.

This rejection process is more effective in giving rise to equi-probable

distribution than the mere modulo reduction technique used in ERNIE. It only requires, for the production of digits, each with a probability $\frac{1}{2}$, that the two noise sources have exactly the same statistical behaviour and this is automatically guaranteed. This follows, since if the probability of a 1 from this noise source is p, the probability of an output is $2pq$ and the probability that this is a 1 is pq. It has the disadvantage that the production of digits is irregular and proceeds at about an average of half the rate if the rejection process were not used. However, since a

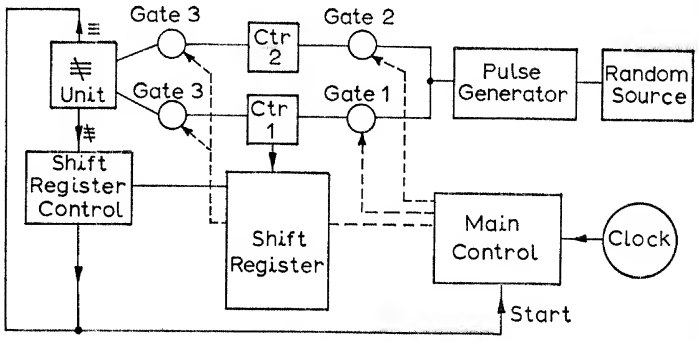


Figure 7

sampling rate at about 20,000 inspections per second is required, the number of pulses or their equivalent which can be counted into the flip-flop is much less by a factor of about 20 than that used in ERNIE. The overall performance of the two machines is probably rather similar.

A mathematical analysis of the rejection techniques shows that not only is the correct probability of each digit produced, but that any residual correlation between successive digits is also reduced. The rejection process has the same advantage as the combination process in ERNIE in that if one part of the mechanism corresponding to one generator fails, the mechanism still produces random digits of quality determined by the elementary process. As in the Manchester Mark I machine, these digits are collected in a register and are available to an external source in groups of 15.

The demonstration that the effect of correlation in the noise sources is reduced in the rejection process requires some assumption of the type of dependence. First suppose that the digits from each source form a Markov chain in which the probabilities of obtaining a 1 after a 0,

or 1, are p_0, or p_1, respectively. The resulting sequence is also clearly Markovian and it only remains to determine the two probabilities corresponding to p_0 and p_1.

Denote the pairs 00, 01, 10, 11 by 0′, 0, 1, 1′ and the probabilities that a sequence of n new symbols terminating with the symbol θ with all other symbols 0′ or 1′ shall follow an old symbol r, by $p_{r,\theta}^{(n)}$. The usual difference equation technique gives

$$p_{r,\theta}^{(n)} = A\alpha^n + B\beta^n$$

where α and β are the roots of

$$(x-q_0^2)(x-p_1^2) - p_0q_0p_1q_1 = 0) \quad , \quad (q_r = 1-p_r)$$

and A and B are constants determined by the initial conditions.

The probabilities that a 1 shall follow the symbol r in the derived sequence are

$$p_r^{(1)} = \sum_{n=1}^{\infty} p_{s,1}^{(n)} = \frac{A\alpha}{1-\alpha} + \frac{B\beta}{1-\beta} = \frac{A\alpha+B\beta-(A+B)\alpha\beta}{(1-\alpha)(1-\beta)} \qquad (s+r=1)$$

Substituting for $A\alpha+B\beta$ and $A+B$ from the initial conditions we obtain

$$p_0^{(1)} = \sum_{n=1}^{\infty} p_{1,1}^{(n)} = \frac{1}{1+q_0+p_1} \quad , \quad p_1^{(1)} = \sum_{n=1}^{\infty} p_{0,1}^{(n)} = \frac{q_0+p_1}{1+q_0+p_1}$$

Write

$$p_1 = p+\delta \quad , \quad p_0 = p-\delta \quad , \quad \delta = \tfrac{1}{2}(p_1-p_0)$$

then

$$p^{(1)} = \tfrac{1}{2} \qquad \delta^{(1)} = \delta/2(1+\delta)$$

δ measures the departure from independence of the digits of the sequences and so the pairing process has reduced the dependence. If the process is repeated n times we obtain

$$\delta^{(n)} = \frac{\delta}{2^n[1+(2-\frac{1}{2}^{n+1})\delta]}$$

It is well known, or easily established, that the probabilities of a Markov chain of n symbols ending in i are $\phi_n^{(i)}$

$$\phi_n^{(0)} = \frac{1}{p_0+q_0}[q_0+p_0(p_1-p_0)^{n-1}]$$

$$\phi_n^{(1)} = \frac{1}{p_0+q_0}[p_0-p_0(p_1-p_0)^{n-1}]$$

Thus the probability of an indefinitely long derived sequence ending in a 1 is

$$p_0^{(1)}/(p_0^{(1)}+q_0^{(1)}) = \tfrac{1}{2}$$

This observation shows how the dependence between the digits of the sequence can be removed entirely after one stage of pairing. Suppose the symbol following each 0 of the derived sequence is changed into its opposite. Then in the Markov scheme $p_0^{(1)}$ is changed into $q_0^{(1)}$. But $q_0^{(1)} = (q_0+p_1)/(1+q_0+p_1) = p_1^{(1)}$. Hence the resulting sequence is an independent sequence. A second pairing will produce a random process.

Next consider a sequence subject to a trend, that is the value of p is steadily increasing or decreasing as the trial proceeds. Let p_r denote the probability of a 1 at the rth trial. Then assuming a linear trend for simplicity, put

$$p_r = p(1+\alpha r) \qquad \alpha \ll p \qquad \alpha r \ll 1$$

Then the probabilities that the rth pair of digits are 01 and 10 are

$$pq+\alpha p^2+2\alpha p(q-\overline{1-\alpha p})r-4\alpha^2p^2r^2$$

and

$$pq-\alpha pq+2\alpha p(q-\overline{1-\alpha p})r-4\alpha^2p^2r^2$$

and the conditional probability that the rth pair of digits will be recorded in the new sequence as a 1 is

$$\frac{pq(1-\alpha)}{2pq+\alpha p(p-q)}\left[1+\frac{2\alpha^2(q-p+\alpha p)}{q(1-\alpha)(2q+\alpha\overline{q-p})}r\ldots\right]$$

$$\sim \tfrac{1}{2}\left(1-\frac{1}{2q}\alpha\ldots\right)\left[1+\frac{\alpha^2(q-p)}{q^2}r\ldots\right]$$

The average number of symbol pairs retained from n digits is of order npq so that the new trend equation is approximately

$$p_r^{(1)} = \tfrac{1}{2}\left[1-\frac{1}{2q}\alpha\right]\left[1+\frac{\alpha^2(q-p)}{2pq^3}r\ldots\right]$$

A repetition of the process gives

$$p_r^{(2)} = \tfrac{1}{2}\left\{1-\frac{\alpha^2(q-p)}{2pq^3}\right\}\left[1+\frac{2\alpha^5(q-p)^2}{p^2q^7}r\ldots\right]$$

Thus the pairing process applied repetitively will rapidly reduce trend.

The pairing process is rather wasteful of digits. Thus, if the sequence happens to be random, on average half the digits are discarded and of

those that are retained each pair gives only a single symbol. The final sequence contains only 25% of the digits of the original sequence. In general, it is the percentage of the information retained in the final sequence that is required. Table 5 shows the quantity for $p = 0{\cdot}1(0{\cdot}1)0{\cdot}5$. The calculations leading to this table have neglected an amount of information contained in the variation of sequence length, but this can be shown to increase as the logarithm of the sequence length, whereas the main contribution increases directly as the sequence length.

This loss of information arises from discarding some of the symbol pairs; we can recover some of this lost information. If the symbol pair 00 is coded as 0 and the pair 11 as 1 in an auxiliary sequence, this will be similar to the original sequence with p replaced by $p^2/(p^2+q^2)$. An application of the pairing process to this sequence will produce further random digits. Repeat these two transformations on the second set of discarded digit pairs and continue in this way indefinitely. The third column of Table 5 gives the total percentage of information retained if this recovery process is used. The efficiency of the process is not much improved. The remaining loss of information arises from the irreversible process involved in transforming the original sequence into the two sequences of accepted and rejected digits.

TABLE 5

p	*Percentage of information in derived sequence*	*Percentage of information after recovery*
0·1	0·19190	0·20250
0·2	0·22163	0·24851
0·3	0·23829	0·28532
0·4	0·24718	0·31584
0·5	0·25000	0·33333

By using more complicated coding of the digits more information can be saved. As an example, suppose the digits of the sequence are taken in sets of 4, and recoded according to the following table:

TABLE 6

Old	*New*	*Old*	*New*	*Old*	*New*	*Old*	*New*
0000	Reject	0100	10	1000	11	1100	Reject
0001	00	0101	01	1001	11	1101	10
0010	01	0110	10	1010	Reject	1110	11
0011	00	0111	00	1011	01	1111	Reject

This is constructed by taking groups of 4 individual configurations with the same probabilities, assigning one member of each set of 4 to the symbol pairs 00, 01, 10, 11 and discarding any spare ones. Applied to a random sequence this process produces a sequence of length three-eighths that of the original sequence. If the rejects 0000 and 1010 are recorded as 0, and 1100 and 1111 as 1, in an auxiliary sequence and this recoded as above and this repeated indefinitely, the final sequence can be extended to a total of 40% of the length of the original sequence. Table 7 shows the percentage of information retained for a general independent sequence by one application of the coding and by its repeated application. Comparison with Table 5 indicates the increase in information content and the comparative stability of the efficiency with p.

TABLE 7

p	*Single application*	*Repeated application*
0·1	0·34926	0·35781
0·2	0·37234	0·38933
0·3	0·37649	0·39903
0·4	0·37571	0·42333
0·5	0·37500	0·40000

If the digits are taken in larger groups and a similar process applied, the efficiency does not increase beyond the values of Table 7. A formal proof that the coding of Table 6 is optimum of its class has not yet been found but exhaustive trials for groups up to 10, the practical limit, show that it is true for that range.

The problem of determining a coding process which retains all the information for any value of p is likely to prove difficult since any coding process to convert such a sequence for a known p to a random sequence involves the consideration of the whole sequence as a unit.

5.7 The General Electric Electronic Probability Generator

This is a machine that has been built for studying non-stationary random walks. This machine does not use the principle of accumulating a large number of pulses in a counter, but each pulse individually is used to give an output digit. It has been built so that the probability of an output digit being one takes the general value p, which is under manual control. A second terminal is given whose output is always

complementary to the main output. It can be likened to an electronic biased penny-tosser.

Figure 8 shows a diagram of this machine.

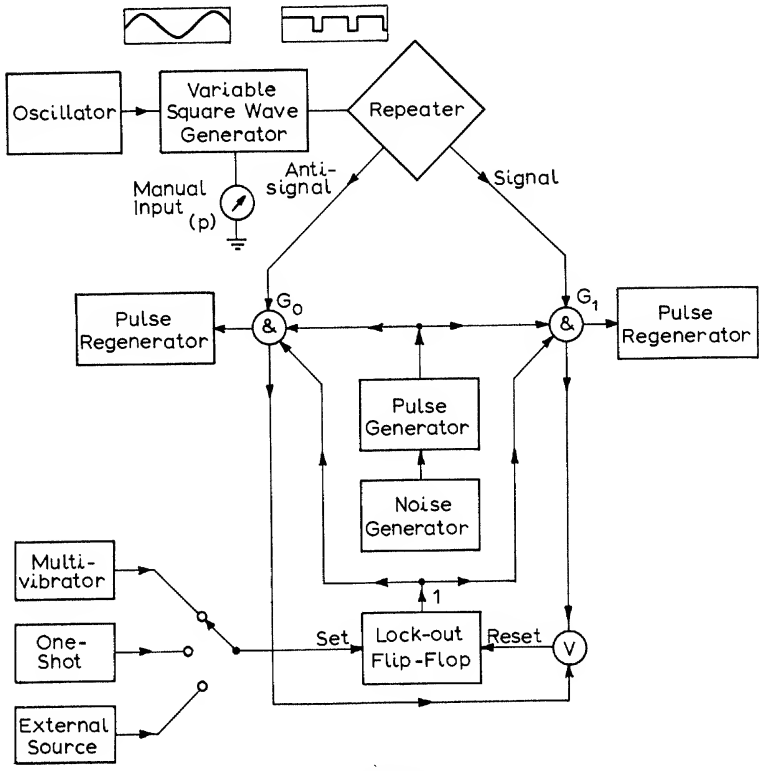


Figure 8

An oscillator drives a series of circuits which produce a regular square wave output in which the proportion p of time that the voltage is high is determined by a manually controllable input voltage. This wave form and its complement are used to control two three-way 'and' gates G_0 and G_1. A second common input to these 'and' gates is from a 'lock-out' flip-flop. When this is set on, the gates are opened to receive pulses from a pulse generator of standard pattern. The first pulse to arrive passes through one, and only one, of the two gates, which gate depending on when the pulse arrives in relation to the square wave input. After passing through a gate the pulse is regenerated and forms

the output on one of the output lines. Thus the probability of it passing the upper gate is determined by the manually set input voltage. In either case, the regenerated pulse is also routed via an 'or' gate to switch the lock-out flip-flop off thereby preventing more than one pulse appearing from the system. The next digit is then read by setting the lock-out flip-flop again. This can be done either by a single shot mechanism, from an external source of any kind, or from a built-in multi-vibrator.

From figures published concerning this machine it is clear that the output of this generator is not of high quality, but it does have a very fast production rate and is no doubt suitable for a wide variety of sampling experiments in which the quality of the random digits is not very important.

The principle of operation is an adaptation of the rotary disc method using two unequal sectors on the disc. The known inferiority of this technique (at least theoretically) may explain the disappointing performance of this rather elegant machine.

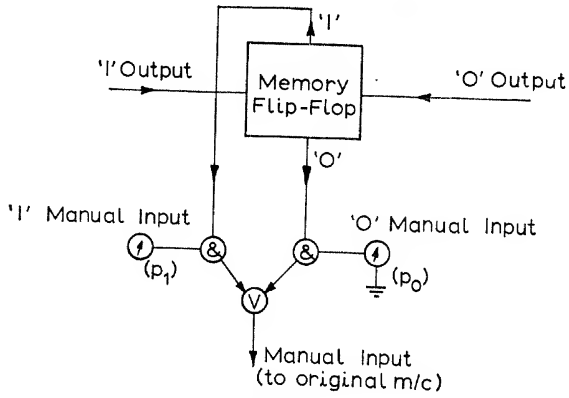


Figure 9

The machine can be further extended to produce dependent binary sequences by means of an additional circuit shown in Figure 9, in which the output digits are used to set a further binary counter which, in turn, controls two gates that determine which of two possible input voltages are applied to the square wave generator. Thus by setting different values for these two voltages, a first order Markov process can be generated.

Although this has not been done in the General Electric machine, it is clear that the process can be generalised and multiple order Markov

chains developed. Figure 10 shows how this is done, for a third order chain.

The conditional probability flip-flop is replaced by a register of such devices, and after each pulse has been obtained on an output terminal the contents of the register are shifted one place, recording in the last vacated place the last digit. A diode matrix based on the digits of this register controls 2^n gates which admit one of 2^n manually set voltages

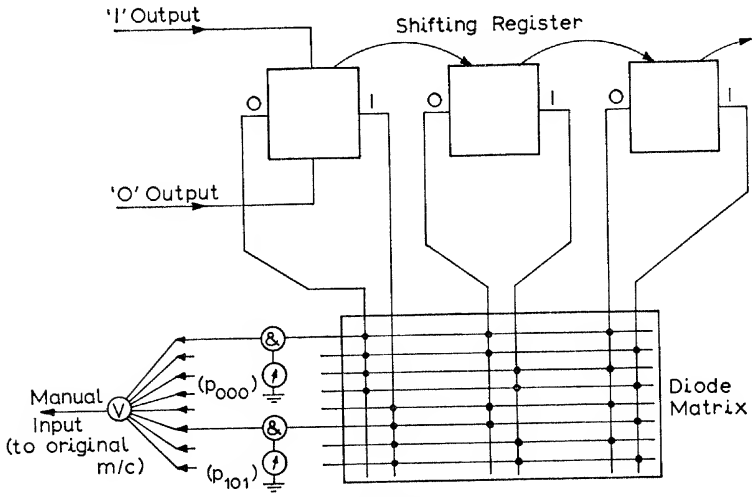


Figure 10

to the square wave generator. Thus, by assigning the right values to these 2^n voltages any n-dimensional Markov chain can be developed.

5.8 The Lion Random Selector

J. R. and K. S. Lion describe a generator using thyratrons (gas-filled triode valves) as the output mechanism, illustrated in Figure 11.

Such a valve will remain non-conducting until the grid is raised to a critical voltage when it 'fires' and will remain conducting. Sinusoidal voltages exactly out of phase are applied to the grids of two thyratrons with a base voltage that ensures that neither thyratron will fire. If the base voltage is now raised, one grid will be raised at that instant above the critical voltage but the other will not. This causes one thyratron to fire and a feedback current is provided, which now ensures that the other valve cannot fire. Which thyratron fires is determined by the phase of

the sinusoid at the moment of raising the base voltage. This is achieved by closing a switch S by a thermal delay and the device depends on the inherent variability of such thermal devices. By choosing a sufficiently high frequency for the oscillation, the variation in the thermal delay changes the grid voltage at a random point in a cycle.

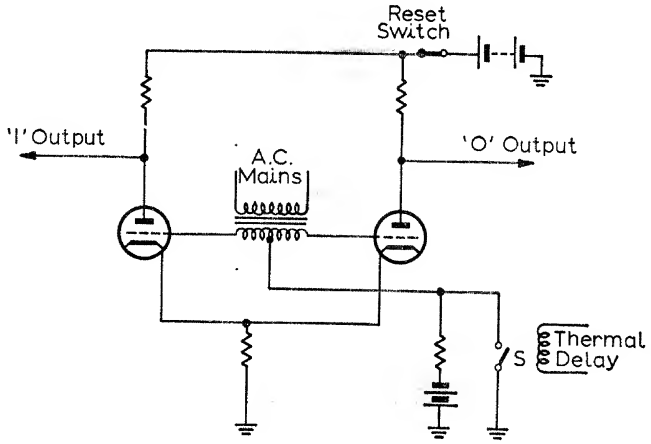


Figure 11

Thus this machine operates on the roulette wheel principle.

The Lion machine uses a 60-cycle oscillation (the American standard mains frequency) and the spread of the thermal delay corresponds to about 3 cycles. Increasing the frequency of oscillation could improve the performance.

Records of individual digit frequencies and of pairs and triplets on a sample of 339 digits gave no indication of departure from randomness.

5.9 The R.C.A. Random Number Generator

This generator utilises the principle of the parmetron, a device used in computing machines developed by the Japanese electrical firm Kanematsu and independently suggested by J. Von Neumann.

The principle of this device depends on the possibility of setting up and maintaining a stable oscillation in an element at a frequency $f/2$ when this is connected to a reactor driven at a frequency f. Two stable configurations are possible exactly out of phase with each other. Thus such an element can be used as a dynamic bistable device.

The advantage of this rather complicated method of achieving two

stable states is the high frequencies that can be used. Several thousand megacycle devices are easily achieved, and very fast computers are in development using this element.

In a computer, the choice of the stable configuration is determined by logical circuits. In the R.C.A. random number generator, the element is set in oscillation from a static state by switching on the main frequency. The phase of the subharmonic oscillation now depends on electrical noise present at the moment of switching.

Thus this is another example of an electronic roulette wheel.

In practice, two 2,000 M/c oscillators are used and the phases of the subharmonic oscillations are compared. If these are equal, a digit one is recorded, otherwise a nought is recorded. At these incredible speeds a sampling rate of about 3×10^7 per second is achieved.

Such high rates of sampling introduce difficulties of testing the output for randomness. The machine described has been tested by counting the number of ones in 228 samples of 1,000 and the resulting distribution checked by a χ^2 test against the theoretical binomial distribution.

5.10 Analogue Randomisers

Random number generators are most suitable for digital applications of the sampling technique. In recent years a large range of analogue computers based on high gain amplifiers with various forms of feedback which give rise to the various linear operators such as adders, differentiators, integrators, etc., have been developed, for the study of analogue processes by machinery. If such devices are assembled to study the effect of random influences on a physical system, a varying source of voltage with a given probability distribution is required.

As far as is known, no machines have been built to provide this facility.

However, various mechanical analogue machines have been built and at least one mechanical analogue randomiser exists. This is the stochastic analogue machine devised by Stafford Beer. The purpose of this machine was to study the behaviour of queues, and to do this, an analogue quantity to represent the process time was required. As one of the main purposes of this machine was to demonstrate the effects of stochastic variation on industrial processes to management, it was felt that the analogy used should be as close as possible and the analogue quantity that was used was time itself.

Thus the problem was reduced to the construction of time intervals, which followed a given probability distribution. This machine used a

mechanical version of the 'Top Hat' method. A ball is dropped on to a sieve which vibrates over a rotating cone and falls into one of 100 pockets at the base of the cone. The oscillations of the sieve and the revolution of the cone result in successive balls falling into the individual pockets with a uniform distribution; it forms a device for selecting one of 100 pockets with equal probability. These pockets can be connected by tubes to merging chambers which, in turn, are connected to another set of pockets. Thus the probability of a ball arriving at one of the merged sockets is proportional to the number of original pockets allocated to that merged group. These groups represent the various possible times in the required distribution, and by adjusting the partitions in the merging chamber, any distribution can be approximated. The merged pockets are arranged horizontally over a moving belt, and the ball(s) prevented from falling on it by a mechanical gate. Simultaneously with the release of a ball into the sieve and cone mechanism, the preceding ball is released from its selected pocket on to the moving belt. The belt is driven at a steady rate, and on reaching the end of the belt, the ball falls down a chute that actuates a micro-switch. The delay in action of this micro-switch is proportional to the distance of the selected pocket from the end of the chain, therefore to the ordinal number of the pocket and thus to the value assigned to that pocket. In this way, a time interval between releasing one ball and collecting another ball constitutes a sample from the specified distribution.

The balls required for dropping into the sieve cone mechanism are held in a zig-zag reservoir and are released on an electrical impulse signal by a solenoid controlled gate. A further unlimited reservoir of balls is provided so that another electrically operated gate may drop a ball into the limited reservoir. This latter represents the queue in front of the process which is itself represented by the analogue machine. The closing of the micro-switch corresponds to the termination of a process on that machine and can be fed to the input of a similar machine to release a ball into the queue before it. The whole machine is illustrated in Figure 12.

A battery of ten of these machines has been built and are used to study complex queueing situations.

5.11 Pseudo-random Number Generators

As we shall see in Chapter 6, there are technical advantages in having a reproducible sequence of numbers instead of purely random numbers. In the same way, in mechanical machines, there may be some advantage

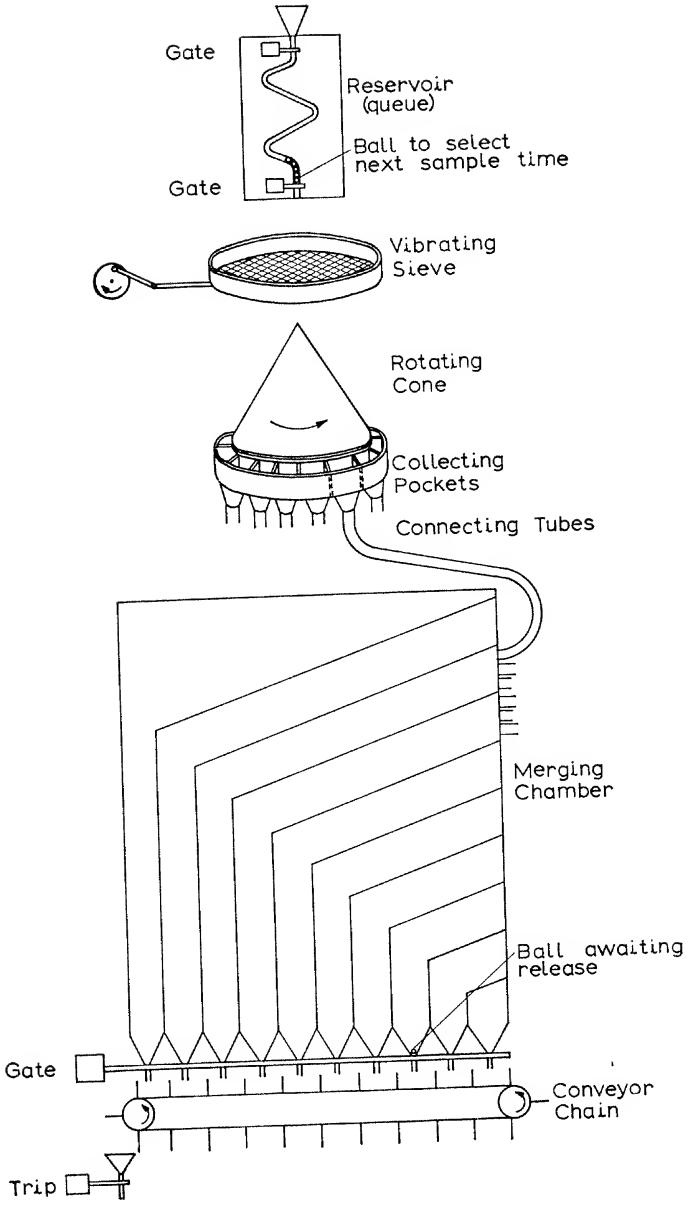

Gate
Reservoir
(queue)
Ball to select
next sample time
Gate
Vibrating
Sieve
Rotating
Cone
Collecting
Pockets
Connecting Tubes
Merging
Chamber
Ball awaiting
release
Gate
Conveyor
Chain
Trip

Figure 12

in having a reproducible sequence of digits that nevertheless display properties similar to random sequences.

Another possible application of such pseudo-random number generators is in demonstration equipment that is designed to illustrate the effects of random behaviour but at the same time is required to be comparatively cheap.

The problem consists of generating a very long cycle of digits. A possible way of achieving a binary sequence of digits that has been used in a queue demonstrator has been constructed at the United Steel Companies.

This machine simulates four process machines, each with an adjustable mean service time, which is served by a common stock of material (a queue) to which items are added at varying intervals of adjustable mean length.

To simplify the machine, each process distribution and the inter-arrival distribution are assumed to be exponential and are defined completely by their mean values.

The machine advances in discrete steps of time marked by a train of equally spaced pulses. At each pulse, the four machines and the arrival point are scanned and a decision made if an event is to occur. This is achieved by inspecting a random binary digit; if this is a one the event occurs, otherwise no event occurs. For the machines, the event is that the present item, if any, being processed is finished; for the arrival point, the event is an arrival.

The assumption of an exponential distribution enables the sequence of digits to be a random one, since this will ensure that the distribution of intervals between events will be a geometric one. This is the discrete analogue of an exponential distribution.

To allow for the varying mean values associated with the different machines, the probability of a one for any given digit must be adjustable. This complication will be ignored in the initial description of the pseudo-random number generator.

The machine is an electro-mechanical one, and the randomising unit is composed of a set of uniselectors. These are electrically operated multi-way multi-pole switches. A bank of 200 contacts, arranged in 8 banks of 25 contacts, are scanned by 8 wipers. On energising the device, all the wipers move from one row of 8 contacts to the next, returning to the first after scanning the last row.

If $2r$ of the contacts randomly selected are earthed and the 200 contacts scanned in some order, then earth will be found on $r\%$ of scans.

If the scan is repeated in the same order, a cyclic pattern of length 200 will develop, but if a large number of different scanning orders can be generated, the cycle length will be much increased. For convenience all the scanning orders will proceed from consecutive rows and only the choice of bank is available to generate different orders.

This is achieved with two further uniselectors. The eight wipers of the first uniselector are wired on to the contacts of the first 23 rows of the second uniselector in a random pattern, so that each occurs 23 times. The eight outlets of this uniselector in turn are each connected to three rows of one bank of the third uniselector in a random order.

To obtain the next digit in the sequence, uniselectors 1 and 2 are stepped and uniselector 3 is stepped only if the last digit is a one. Uniselectors 2 and 3 skip over their blank levels.

The operation of this mechanism has defied all attempts at theoretical analysis of its behaviour, but has been found in practice to generate a sequence of digits that pass all the classical tests for randomness.

An indication of the length of cycle obtainable can be obtained as follows.

If the contacts of the first uniselector are numbered in any conventional way, the number of the contact connected to the output is defined by the levels r, s, t of the three uniselectors and so the presence or absence of earth can be represented by a function $u(r, s, t)$, taking values 0 or 1 according to the earth condition of the connected contact.

Now denoting nth values by a sub-script, we have

$$\begin{aligned} r_{n+1} &= (r_n+1) \bmod 25 \\ s_{n+1} &= (s_n+1) \bmod 23 \\ t_{n+1} &= (t_n+u_n) \bmod 24 \end{aligned}$$

Assuming r, s, t start at zero, the cycle length is given by the least n for which

$$r_n = s_n = t_n = 0$$

i.e. $n \equiv 0 \bmod 25$

$n \equiv 0 \bmod 23$ i.e. $n \equiv 0 \bmod 575$

$$U_n = \sum_{r=0}^{n-1} u_r \equiv 0 \bmod 24$$

Now assume the mechanism is perfect; then Σu_r is a binomial variable with probability p and index n.

The distribution U_n will be approximately uniform and thus the probability that the cycle is of length $575m$ is

$$p(m) = \tfrac{1}{24}(\tfrac{23}{24})^{m-1}$$

and the expected cycle length is $24 \times 575 = 13{,}800$.

As this crude theory predicts, the cycle length does depend on the starting position. The theory, of course, breaks down as the distribution of U_n cannot be of the form assumed since there is a limited amount of variability in the machine represented by the initial wiring.

In practice, the contacts on uniselector 1 are divided into 20 groups of 10 randomly disposed, and r of these are earthed if the probability of a one is required as $5r\%$. Switches associated with each machine and with the arrival point are used to determine the groups used for each type of event and are chosen differently for each type.

Thus the stepping of uniselector 3 depends on the last use made of the mechanism. This further complication seems to lengthen the cycle although the crude analysis alone would not predict this.

CHAPTER 6

PSEUDO-RANDOM NUMBERS

As the complexity of the sampling experiments performed increases, the possibility of computing errors increases. To repeat the calculations will require the random numbers used in the initial calculation to be available for the check calculation. Using tables of numbers and manual methods of calculation, this is possible, but tedious.

If automatic calculation is used, the storage of the large volume of random numbers used becomes a serious problem. It is this storage problem that has led to the abandonment of the use of tables of random numbers in computers. The use of the physical generators described in the last chapter will not eliminate this storage problem if check calculations are required.

We shall see later that the ability to repeat the sequence of the random numbers has valuable advantages in increasing the accuracy of sampling experiments.

Thus the need arises for a process that will generate a 'random number sequence' which will be reproducible. This is clearly a complete contradiction of the meaning of a random sequence. The reproducibility of the sequence implies the possibility of prediction of the sequence and constitutes the proof of its non-randomness.

However, the advantages of reproducibility demand an investigation if it is possible to obtain approximately random sequences by a deterministic process. In principle it should be possible to define certain criteria required of a sequence that is to be used as a random sequence and then construct a process that will satisfy the criteria. Such a sequence is called pseudo-random.

An obvious criterion required is that each element of the sequence is bounded and that all possible values within the bounds will appear equally often in the sequence. This is clearly insufficient since the sequence

$$1, 2, 3, \ldots, n \qquad 1, 2, 3, \ldots, n \qquad 1, 2, \ldots, n$$

satisfies this requirement and yet is clearly not random.

A second criterion required might be that the serial correlation of consecutive elements in the sequence is zero. However, the sequence

$$1, 2, 3, \ldots, 2n \qquad 1, 2n, 2, 2n-1, 3, 2n-2, \ldots, n$$
$$n+1, 1, 2, 3, \ldots, 2n, \ldots$$

will have this property for large n and yet it is clearly not random.

We may extend the requirements by demanding higher and higher serial correlations to be zero or more strongly that longer and longer sequences are equally likely.

Again, this will not prevent an obviously non-random sequence being accepted as pseudo-random. The sequence

$$1, 2, 2, 3, 3, 1, 3, 2, 1, 1, 2, 2, 3, 3, 2, 3, 2, 1, \ldots$$

contains each of the 9 possible ordered pairs (1, 1), (1, 2), (1, 3), (2,1), (2, 2), (2, 3), (3, 1), (3, 2), (3, 3) equally often but is clearly cyclic.

All deterministic rules for forming a sequence of bounded numbers will in fact be cyclic and the essential problem is to produce a sufficiently long cycle.

The first attempts to generate pseudo-random sequences were based on arbitrary rules and the properties of the resulting sequences could not be predicted theoretically. The value of a rule was determined by applying the standard tests for random numbers to the sequences obtained.

6.1 The Mid-square Technique

This was first introduced by Von Neumann. A p-digit number x_0 is squared and from the resulting $2p$ digits the mid digits are taken as x_1. The number x_1 is then squared and the process repeated. As an example take $p = 2$ and

$$x_0 = 76 \quad \text{then} \quad x_0^2 = 5776 \qquad x_1 = 77$$

The sequence is

$$76, 77, 92, 46, 11, 12, 14, 19, 36, 29, \ldots$$

If the radix used is r then there are r^{2p} possible values of x and consequently the sequence must ultimately repeat some previous value. From that point the sequence is repeated—i.e. it is cyclic. In practice, the cycle length is considerably less than the theoretical maximum r^{2p}. The length of the cycle is dependent on the starting value x_0. Certain values can lead to a zero term when the cycle length becomes one.

These objections are sufficient to condemn this method, but a more serious objection is that the sequences obtained do not satisfy the primary requirement of a random sequence that any value in the permitted range is equally likely to occur.

This can be shown as follows.

The truncation of the digits from the back of x_n^2 (in forming x_{n+1}) is merely to limit the number of digits that are handled and in a theoretical analysis we merely consider the front truncation.

Then

$$x_{n+1} = r^p(x_n^2 \bmod r^{-p})$$

Put $\quad x_n^2 = u \qquad r^{-p} = h \qquad m = r^p = \dfrac{1}{h} \qquad x_{n+1} = \omega$

Then

$$mh = 1$$

If u and v have p.d.f.'s $p(u)$ and $p_1(v)$ and c.d.f.'s $P(u)$ and $P_1(v)$ respectively, then

$$p_1(\omega) = h[p(\omega) + p(\omega+h) + p(\omega+2h) \ldots p(\omega+\overline{m-1}h)]$$

$$\sim \int_0^{(m-1)h} p(\omega+x)\,dx = P(\omega+1-h) - P(\omega) \quad (1)$$

Now if x_n has a uniform distribution, we have

$$p(u) = \tfrac{1}{2}u^{-\frac{1}{2}} \quad , \quad P(u) = \sqrt{u}$$

$$P_1(\omega) \sim \tfrac{1}{2}(1-h) - \tfrac{1}{8}\frac{(1-h)^2}{\sqrt{\omega}} + 0\left(\frac{1}{\omega^2}\right)$$

$$P_1(u) = u + \sqrt{n}(\sqrt{u} + 0(u)) + 0(hu)$$

Thus x_{n+1} has a skew distribution favouring low values of x_{n+1}.

This bias in x_{n+1} will create a greater bias in x_{n+2} and will cause a continuous degradation of the sequence.

6.2 The Mid-product Method

In an attempt to remedy this defect, an extension of the mid-square method has been proposed. Two starting values x_1, x_2 are given and their product u is formed. The mid digits of u are used as x_3. The process is now repeated on x_2, x_3 to form x_4 and so on.

The distribution of u (assuming a uniform distribution on (0, 1) for u_1 and u_2) is

$$p(u) = -\log u$$

Applying formula (1) we obtain

$$p_1(u) = u - hu \log u + 0(hu)$$

This process has less bias than the mid-square method and due to the need for a pair of terms to repeat for a cycle to restart usually has a larger cycle than that method.

However, these two processes do not now receive much favour because of the known bias. Various other transformations followed by truncation could be adopted, but all will lead to bias. Within the approximation of (1) we require to determine a transformation $f(x_n)$ of x_n so that x_{n+1} given by

$$u = f(x_n)$$
$$x_{n+1} = r^p(u \bmod r^{-p})$$

will have a uniform distribution.

Within the limits of the approximation (1) this implies

$$P(\omega + 1 - h) - P(\omega) = A$$

for a suitable range of ω and some constant A.

Over this range $P(\omega) = \alpha + \beta\omega$ for suitable α and β and hence

$$p(u) = \beta$$

Thus the distribution of u is uniform and so the only permissible transformation is

$$f(x) = kx + l \qquad (k, l \text{ constants})$$

The case $l = 0$ has been suggested for completely different reasons by Lehmer.

6.3 The Lehmer Congruence Method

Once it is recognised that a pseudo-random sequence will be cyclic, it is possible to require to generate a cycle of maximum possible length. In 1949, Lehmer made the suggestion to use the theory of numbers to devise such long cyclic sequences.

The theory of congruences supplies such sequences. Consider an integer sequence suggested by the heuristic argument of the last section

$$x_{n+1} = kx_n \bmod m$$

where initially m and k are arbitrary integers.

Then

$$x_n = k^n x_0 \bmod m$$

For the cycle length we require to find the minimum n, for which

$$k^n \equiv 1 \pmod{m} \qquad (2)$$

$$(\text{assuming } (x_0, m) = 1)$$

Now the Fermat-Euler Theorem states that if $(k, m) = 1$ then

$$k^{\phi(m)} \equiv 1 \pmod{m}$$

where $\phi(m)$ is the number of integers less than and prime to m. If m is prime, $\phi(m) = m-1$ and then it is apparently possible to achieve a sequence of length $m-1$.

However, the theorem does not state that $\phi(m)$ is the smallest integer for which (2) is true. It is clear that the smallest integer n with the property (2) is a divisor of $\phi(m)$, for otherwise we can write

$$\phi(m) = an+b \quad , \quad 0 < b < n$$

Then

$$k^{\phi(m)} \equiv k^{an}k^b \equiv k^b \equiv \text{mod } m$$

i.e. b satisfies (2)
in contradiction of the definition of n as the least value with property (2).

It remains to choose k and m to make n as large as possible. If the maximum length cycle can be achieved the starting value x_0 will be immaterial, and if n cannot be made equal to $m-1$ it is still desirable to make the cycle length independent of x_0.

However, there is a prior consideration in the choice of m. The operation of taking a congruence involves a division and it is desirable to choose m to make this usually lengthy operation rapid.

For a number in radix scale r choose an integer p and write

$$x = yr^p+z = y(r^p-1)+y+z$$

Then

$$x \pmod{r^p} = z$$

$$x \pmod{(r^p-1)} = y+z$$

Thus if m takes either of the special values r^p, r^p-1 the division merely involves the separation of the digits of the number into two parts and at most the addition of the parts. Also since

$$x = y(r^p+1)+z-y \equiv (y-1)(r^p+1)+(r^p-y)+z+1$$

$$x \pmod{(r^p+1)} = z-y \qquad z \geqq y$$

$$= r^p+1+z-y \qquad z < y$$

Thus r^p+1 is also a suitable divisor to eliminate any real work in division.

If
$$m = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$$
where $p_1, p_2 \dots$ are primes and $a_1, a_2 \dots$ are positive integers, then
$$\phi(m) = p_1^{a_1-1}(p_1-1)p_2^{a_2-1}(p_2-1)\dots p_k^{a_k-1}(p_k-1)\dots \qquad (3)$$

With most computers working in the binary scale, $r = 2$ and we are interested in $\phi(2^p)$, $\phi(2^p-1)$ and $\phi(2^p+1)$. From (3) it follows that $\phi(2^p) = 2^{p-1}$.

It is possible to show that in fact for $m = 2^p$ the maximum cycle is of length 2^{p-2} and the only value of k giving this cycle must satisfy the conditions
$$k \equiv \pm 3 \bmod 8$$
$$x_0 \text{ odd}$$

A suitable choice of k is an odd power of 5 since
$$5^{2q+1} \equiv (1+4)^{2q+1} \equiv (1+(2q+1).4) \bmod 8 \equiv -3 \bmod 8$$

In this case, p should be chosen as large as possible and hence equal to the number of digits in the number stored. After multiplication by k the least significant p digits of the double length product are used as x_{n+1}.

Values for this scheme which have been quoted in the literature are

$k = 5^{13}$	$p = 36, 39$
$k = 5^{17}$	$p = 40, 42, 43$

For decimal machines we are concerned with
$$\phi(10^p) = \phi(5^p)\phi(2^p) = 2^{p+1}\,5^{p-1}$$

There is little in the literature concerning the theory for this scheme but the following values have been used

$k = 7$	$p = 10$	
$= 7^{4a+1}$	$p = 11$	$a \geqq 0$
$= 3^{19}$	$p = 20$	

The case $m = 2^p-1$ is more complex. The longest cycle will be given when p is chosen so that 2^p-1 is a prime when the cycle length is 2^p-2.

Now 2^p-1 is prime only if p is prime since
$$(x^{pq}-1) = (x^p-1)\{x^{p(q-1)}+x^{p(q-2)}\dots+1\}$$

The numbers 2^p-1 for prime p are known as Mersenne numbers and it has been established that the only prime Mersenne numbers are given by

$$p = 2, 3, 5, 7, 13, 17, 19, 31, 61 \ldots$$

Thus for this case longer cycles may be achieved by using a value of p less than the word length. The practical values of p used are 19 and 31 giving cycle lengths of 524, 286 and 147, 483, 646 respectively.

In the case of $m = 2^p-1$, the choice of a suitable value for k is not amenable to a complete analytic solution. A process of trial and error for small k is usually used. Any power of such a k can also be used providing it is prime to the period $m-1$; also a value congruent mod m to such a power will also give the maximum period.

The commonest combination is that given by $p = 31$, $m = 2^{31}-1$, $k = 13^{13} \equiv 455{,}470{,}314 \bmod (2^{31}-1)$.

$$x_{n+1} = 455{,}470{,}314x_n \bmod (2^{31}-1)$$

Any value in the range $1 \leqq x_0 \leqq 2{,}147{,}483{,}646$ may be used for starting the sequence.

This sequence is usually the best for fixed-point binary machines.

The case $m = 2^p+1$ is not usually so satisfactory. This is only prime if $p = 2^q$ for some integer q. The numbers $2^{2^q}+1$ are known as Fermat's numbers and the first four, 5, 17, 257, 65,537, are the only known prime values in the sequence in spite of investigation at least as far as $q = 73$.

Thus composite m must be used and the one useful case is given by $p = 29$.

$$m = 2^{29}+1 \qquad \phi(m) = 3, 033, 168$$

and

$$k = 7^{11} \equiv 366{,}714{,}004$$

$$x_{n+1} = 366{,}714{,}004x_n \bmod (2^{29}+1)$$

This is useful in a floating machine (such as the Ferranti Mercury) where the mantissa is restricted to about 30 binary digits.

An alternative sequence sometimes used is

$$p = 35$$
$$m = 2^{35}+1 \qquad \phi(m) = 1, 034, 040$$
$$k = 23$$
$$x_{n+1} = 23x_n \bmod (2^{35}+1)$$

For decimal machines, values recommended are

$$p = 8 \qquad m = 10^8+1 \qquad \phi(m) = 5, 882, 352$$
$$k = 23$$

The heuristic argument of the last section suggests that the most general transformation that would give a usable cycle is

$$x_{n+1} = (kx_n + l) \pmod{m}$$

If $l \neq 0$, it is possible for the sequence to include the value zero, and a cycle of length m is possible.

If $m = 2^p$ a cycle of length 2^{p-1} can be produced by the recursions

$$x_{n+1} = (4k+1)x_n + k \pmod{2^p}$$

and

$$x_{n+1} = (4k+1)x_n + 3k \pmod{2^p}$$

provided k is odd.

These schemes all have the property that each of the permitted numbers appears once and once only in each cycle. This however does not indicate how the values are distributed in the cycle. The choice of a large value for k is to prevent a small value of x_n being followed by the small values kx_n, $k^2x_n \ldots$. A large value of k causes kx_n to exceed m and so x_{n+1} is not necessarily small.

Some indication of the distribution of the values in the sequence is given by the serial correlation of the sequence. If the parameters k and l are chosen so that the cycle is of full length m, it is possible to estimate this.

$$\rho(x_n, x_{n+1}) = \left\{\sum_{n=0}^{m-1} (x_n x_{n+1}) - \frac{1}{m}\left(\sum_{n=0}^{m-1} x_n\right)^2\right\} \Big/ \left\{\sum_{n=0}^{m-1} x_n^2 - \frac{1}{m}\left(\sum_{n=0}^{m-1} x_n\right)^2\right\}$$

Now

$$\sum_{n=0}^{m-1} x_n = \sum_{r=0}^{m-1} r = \tfrac{1}{2}m(m-1)$$

$$\sum_{n=0}^{m-1} x_n^2 = \sum_{r=0}^{m-1} r^2 = \tfrac{1}{6}m(m-1)(2m-1)$$

since the sequence $\{x_r\}$ is merely a rearrangement of the sequence $0, 1, \ldots, m-1$.

It remains to evaluate $\sum_{n=0}^{m-1} x_n x_{n+1}$

Put

$$kx_n + l = q_n m + r_n$$

where

$$0 \leqq r_n < m$$

Then we have

$$x_{n+1} = r_n$$

$$\Sigma x_n x_{n+1} = \Sigma x_n (kx_n + l - q_n m) = k\Sigma x_n^2 + l\Sigma x_n - m\Sigma x_n q_n$$

Thus the problem reduces to evaluating $\Sigma x_n q_n$.

Now if $q_n = s$ then x_n satisfies the inequality.

$$0 \leqq kx+l-sm < m$$

$$\frac{sm-l}{k} \leqq x < \frac{sm-l+m}{k}$$

i.e. there are approximately $\frac{m}{k}$ $(=\mu$ say) consecutive values of x for each s. The possible values of s are $0, 1, \ldots, k$. The highest value k only occurs once at the end of the cycle. Neglecting this we have

$$\Sigma xq \simeq \sum_{s=0}^{k-1} s \sum_{t=0}^{\mu-1} (s\mu+t) = \mu^2 \sum_{s=0}^{k-1} s^2 + \frac{\mu(\mu-1)}{2} \sum_{s=0}^{k-1} s$$

From this an approximation to ρ can be found, neglecting terms in $\frac{1}{m}$ as

$$\rho(x_n, x_{n+1}) = \frac{1}{k} - \frac{6l}{km}$$

This is only valid for small l.

A more accurate formula involves the least positive residues r_q of the numbers $l-qm$.

Let $$S = \sum_{q=0}^{k} r_q q$$

Then if $l < k$

$$\rho(x_n, x_{n+1}) \simeq \frac{1}{k} + \frac{12}{m}\left(\frac{S}{k^2} - \frac{k}{4}\right)$$

$\geqq k$

$$\rho(x_n, x_{n+1}) \simeq \frac{1}{k} - \frac{6l}{mk}\left(1 - \frac{l}{m}\right) + \frac{12}{m}\left(\frac{S}{k^2} - \frac{k}{4}\right)$$

Greenberger, who is responsible for the later formula, has developed techniques for determining S for any given generator and, for example, has shown that for $m = 2^{35}, k = 2^{34}+1, l = 1, \rho = \frac{1}{4}$. This shows that the generation of a long cycle is no guarantee that the sequence will be suitable as a generator.

Another example quoted by Greenberger is more satisfactory. Again

$$m = 2^{35} \qquad k = 2^{18}+1 \qquad l = 1 \qquad \rho \ll 2^{-18}$$

The analysis will also give higher serial correlations by noting that x_{n+m} is given by

$$x_{n+m} = \left[k^m x_n + \frac{(k^m-1)l}{(k-1)}\right] \bmod m$$

If this generator also gives a complete cycle, then the formula can be applied with k replaced by k^m and l by $\frac{l(k^m-1)}{k-1}$.

This interesting development has not yet been extended to cases in which a full cycle is not generated.

However, even if the correlational properties of the sequences are satisfactory, the main objection to them for sampling purposes is that once a value has occurred in the sequence, it cannot occur again in the present sampling experiment. This denies the principle of independence of the items in the sample drawn.

In practice, this is readily overcome by truncating each number in the sequence, removing, say, the last 7 digits. Now a truncated value can occur exactly 128 times in a cycle and the number of times it will appear in a short stretch of the cycle will vary. The appearance of independence has been restored.

6.4 Second-order Recurrence Processes

The mid-product method produces a larger cycle than the mid square in general, and this raises the question if a second order process

$$x_{n+1} = f(x_n, x_{n-1}) \bmod m$$

will produce a longer sequence. The simplest possibility is

$$f(x_n, x_{n-1}) = x_n + x_{n-1}$$

$$x_{n+1} = (x_n + x_{n-1}) \pmod m$$

Any sequence defined by

$$x_{n+1} = x_n + x_{n-1}$$

is known as a Fibonacci sequence and has many interesting properties. The most well-known sequence is given by $x_0 = x_1 = 1$ and is given by the formula

$$u_n = \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right\}$$

The general sequence is

$$x_n = u_{n-1}x_0 + u_n x_1$$

The cycle length is given by the least n satisfying

$$(u_{n-1}-1)x_0 + u_n x_1 \equiv 0 \bmod m$$

$$u_n x_0 + (u_{n+1}-1)x_0 \equiv 0 \bmod m$$

which reduces to

$$u_n^2 - (u_{n+1}-1)(u_{n-1}-1) \equiv 0 \bmod m$$

since at least one of x_2, x_1 is non-zero.

Now it can be shown that

$$u_n^2 - u_{n+1}u_{n-1} = (-1)^{n-1}$$

and the condition reduces to

$$v_n = 2u_{n-1} + u_n \equiv (-1)^{n-1} - 1 \pmod m$$

Now for odd n

$$2^{m-1}v_n = 2[(1+\sqrt{5})^m + (1-\sqrt{5})^m] = 1 + {}^mC_2 5 + {}^mC_4 5^2 \ldots m5^{\frac{1}{2}(m-1)}$$
$$\equiv 1 \bmod m$$

since ${}^mC_r \equiv 0 \bmod m$

By the Euler-Fermat theorem, if p is prime, $2^{p-1} \equiv 1 \bmod p$ and thus for prime p the condition becomes

$$v_p \equiv 1 \bmod p$$

For m an odd prime, the maximum cycle of length m cannot be achieved. The complete theory involves the theory of primitive roots of congruence equations which is beyond the scope of this book.

The additive congruence method has not found much favour. However, Page has pointed out that since

$$x_n = \frac{1}{\sqrt{5}}\left[\left(x_0 + \frac{1+\sqrt{5}}{2}x_1\right)\left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \left(x_0 + \frac{1-\sqrt{5}}{2}x_1\right)\left(\frac{1-\sqrt{5}}{2}\right)^{n-1}\right]$$

$$x_n \simeq Ax^n \quad , \quad A = \left(\frac{2}{5+\sqrt{5}}\right)\left(x_0 + \frac{1+\sqrt{5}}{2}x_1\right) \quad , \quad x = \frac{1+\sqrt{5}}{2} \quad ,$$

x_n behaves like the solution of $x_{n+1} = Ax_n$, i.e. asymptotically the method behaves like the multiplicative congruence method with an irrational choice of constant multiplier.

If the pseudo-random numbers are used as a source of random digits rather than as approximations to uniform random variables some care is necessary. If congruences are taken with respect to 2^m, clearly defined patterns are created in the trailing digits.

For the multiplication congruence method

$$x_{n+1} = kx_n \bmod 2^m$$

$$y_{n+1} = x_{n+1} \bmod 8 = ly_n \bmod 8 \quad \text{where} \quad l \equiv k \bmod 8$$

For a long cycle $\quad l = \pm 3 \qquad x_0 \equiv 1 \bmod 2$

$$l^2 \equiv 1$$

Thus the last three binary digits can only take two values, y_0 and $3y_0 \pmod 8$ or y_0 and $5y_0 \pmod 8$, i.e. two digits have fixed values and the other alternates from term to term.

By taking congruence to 2^p a more general result can be obtained.

Put $k' \equiv k \bmod 2^p$. Let a be least positive solution of

$$k'a \equiv 1 \bmod 2^p \quad , \quad y_n \equiv x_n \bmod 2^p \quad , \quad y_{n+1} = k'y_n \bmod 2^p$$

The only values of y_n are $y_0, k'y_0, k'^2y_0, \ldots, k'^a y_0$. Thus the last p digits are cyclic of order 2^{p-1}.

Similarly with the additive congruence method if $y_n \equiv x_n \bmod 2^p$

$$y_{n+1} = (y_n + y_{n-1}) \bmod 2^p$$

and y_{n+1} will be cyclic of period at most 2^p. We have shown that for odd prime p the period will be less than 2^p.

If congruences are taken to an odd modulus this phenomenon largely disappears as the congruences to the composite modulus $2^p m$ will give a more complex structure.

However, the last digit in the multiplicative process will be alternately odd and even if k is odd and in the additive process it is cyclic of period three (even, even, odd).

If random digits are required from such pseudo-random numbers a 'purifying' process such as the rejection technique described in the chapter on physical random generation is advised.

6.5 Tests on Pseudo-random Numbers

In spite of the doubts expressed in many quarters of the validity of these numbers as a source of randomness, there are surprisingly few published accounts of the results of applying the standard tests to numbers generated in this way.

The first of these, by Johnson, concerned the multiplicative congruence method with

$$k = 23 \qquad m = 2^{35}+1 \qquad x_0 = 10,987,054,321$$

A sample of 112,000 35-bit numbers were paired and the resulting 70-bit numbers divided into 7 10-bit numbers. The resulting 392,000 numbers were divided into 28 blocks of 14,000 numbers and each block subjected to three tests:

(i) the distribution of 1,024 distinct numbers against the theoretical uniform distribution;
(ii) the number of 1's in the 140,000 digits;
(iii) the poker test.

The resulting χ^2 for each block were then tested by a χ^2 test for agreement with the theoretical χ^2 distribution. On all three tests, the sample passed this composite test.

Thus the evidence of this experiment is that this particular sample behaved under test similarly to real random numbers.

A similar investigation has been made for the multiplicative congruence method $k = 13^{13}$, $m = 2^{31}-1$. No details have been published but the conclusion was the same as in the preceding case.

The third account, by Davis and Rabinowitz, concerned the use of random numbers to pick a random point in an n-dimensional cube.

The test used was that the correct proportion of such points lay inside the inscribed n-dimensional sphere.

Three methods of generating pseudo-random numbers were used:

(i) the multiplicative method,

$$k = 5^{17} \qquad m = 2^{42} \qquad x_0 = 1$$

(ii) the additive method, $m = 2^{44}$, $x_0 = 0$, $x_1 = 1$, selecting every other value generated;
(iii) the same method, skipping 1, 2 and 3 terms depending upon whether the last hexi-decimal digit of the last number selected was 1, 2 or 3 (mod 3).

Starting values $x_0 = [2^{42}.\pi]$ $x_1 = 5^{17}$

The conclusions of this test are that methods (i) and (iii) are satisfactory, but not method (ii).

CHAPTER 7

ELEMENTARY SAMPLING EXPERIMENTS

The original sampling experiments, conducted by the early statisticians such as 'Student', Galton, Karl Pearson and others of his school, were concerned with determining the form of the distribution of statistics. The object of these studies was to determine the amount and form of variability that some estimate of a parameter of a basic distribution would have in repeated sampling. With this knowledge, it was possible to make statements concerning the likeliness that a particular sample obtained from some experiment could have arisen from some hypothetical distribution.

As an illustration of this process, consider an estimate for the mean of a normal distribution, given the sample of the n values of such a distribution. What estimate of the mean should be used, and how will it be distributed in repeated samples of n? In fact, as is well known, the average of the sample values is the best possible estimate of the mean and its distribution can be shown by analytic methods to be normal. If the standard deviation of the original distribution is σ then that of the mean is $\frac{\sigma}{\sqrt{n}}$.

Now suppose the particular sample of say 16 values from such a normal distribution has an average of 1·29 and it is known that the standard deviation is 2·0. The standard deviation distribution of the mean of repeated samples of 16 values will be 0·5. If we require to test if the original distribution has a zero mean, we ask if it is likely that such an extreme value as 1·29 could arise from a normal distribution of mean zero and standard deviation 0·5. Consulting tables of the normal probability interval we find that

$$Pr(|x| \geqq 1{\cdot}29) = 0{\cdot}01 \tag{1}$$

i.e. only once in a hundred samples will a value as extreme as this from the expected occur.

The assumption of a zero mean is unlikely. It is not, of course, disproven, but in any situation in which indefinitely large departures from the expected values can occur, no hypothesis will be proved or dis-

proved, but will only be shown to be likely or unlikely on the evidence of the sample given.

It was arguments of this kind that led to the modern theory of significance testing, then the Neyman-Pearson theory of hypothesis testing, and in modern times to statistical decision function theory.

The simple example above made an assumption unlikely to be realised in practice: a question is asked about the mean of a distribution when its standard deviation, the measure of spread, is assumed known. What if the spread is also unknown? This will also have to be estimated from the data. A natural estimate for the expected mean square deviation will be a sample mean of the squares of the deviations from the sample mean. In symbols, if the sample values are

$$x_1, x_2, x_3, \ldots, x_n$$

then

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

In practice, it is easier to deal with ns^2 rather than s. It is possible to obtain the theoretical distribution of the former quantity by mathematical analysis. This is the well-known χ^2 distribution and its degrees of freedom are $n-1$. Thus we have

$$E(ns^2) = (n-1)\sigma^2$$

and so an unbiased estimate of σ^2 is given by $ns^2/(n-1)$.

Now if a given sample of 16 values from a normal distribution of unknown mean and standard deviation has a mean of value 1·29 an estimate of σ is required before the probability statement (1) can be made, but this estimate is liable to error. The result of substituting a wrong value of σ will lead to a false value of the probability and to a false conclusion. How is this dilemma to be resolved?

In 1908, 'Student' recognised that the correct approach to this problem is to form a statistic

$$x = \frac{\bar{x} - \mu}{\sqrt{n}s} \tag{2}$$

and to determine its sampling distribution. He actually derived the analytic expression for this distribution and this achievement is often regarded as the foundation step of modern mathematical statistics.

Whether, from doubt of the validity of this argument, or for the purposes of convincing the readers of his result, he also established the distribution by a sampling experiment. Since the number of samples that can be drawn will be limited by time and energy, the resulting theoretical distribution cannot be reproduced exactly. In practice, a histogram is found, and the shape of this histogram is compared with the theoretical distribution. Suppose we are concerned with a sample size of 16, then samples of 16 observations from a normal distribution of mean zero and standard deviation 1 are drawn. From these, the statistic t given by (2) is calculated. This is repeated, say, 500 times and the resulting 500 values of t are formed into a histogram. The choice of the number of class intervals and their width can be chosen for the histogram by inspection of the resulting values of t. The results of such an experiment are given in Table 8.

TABLE 8

t	f	t	f
0·0–	37		
0·2–	36	—0·2	39
0·4–	40	—0·4	41
0·6–	24	—0·6	32
0·8–	28	—0·8	39
1·0–	25	—1·0	19
1·2–	12	—1·2	17
1·4–	19	—1·4	14
1·6–	10	—1·6	8
1·8–	7	—1·8	11
2·0–	9	—2·0	3
2·2–	5	—2·2	8
2·4–	3	—2·4	3
2·6–	1	—2·6	5
2·8–	2	—2·8	3

If the resulting histogram appears to be too ragged, then the choice lies between choosing wider class intervals, thereby sacrificing definition of the distribution or taking further samples. It is very desirable to have some objective method of obtaining an estimate of the sample size required before the sampling starts. This must be chosen to give a prescribed accuracy to the estimates of some frequency class. Supposing the true probability of falling in a particular class has a value p. Then given sample size n, the number of sample values falling in this class will have a binomial distribution of mean np and variance npq. If n is

large enough, as it will usually be in practical experiments, the distribution may be approximated by a normal distribution and thus the probability of any arbitrary number falling in the cell can be easily determined. However, it is not the distribution of the absolute number of elements falling in the cell that is required, but the accuracy in which the class frequency can be estimated. The relative error in estimating the class frequency is given by $(x-np)/np$ where x is the number appearing in the class. This has an approximately normal distribution with mean zero and standard deviation $\sqrt{q/np}$. Thus to find the probability P of maintaining a given level of accuracy α we have to determine n from the equation

$$P = \text{Prob}\left(\left|\frac{x-np}{np}\right| \leqq \alpha\right)$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-k\alpha}^{k\alpha} e^{-\frac{1}{2}t^2}\,dt = \Phi(k\alpha)$$

where

$$k = \sqrt{\frac{np}{q}}$$

This can be written in the form

$$k = \frac{1}{\alpha}\Phi^{-1}(P) = K \quad \text{say}$$

or

$$n = K^2.\frac{1-p}{p} \tag{3}$$

This is shown for suitable values of K in Figure 13 and illustrates that the smaller the class frequency we try to estimate the less accurate is our estimate. The sample size n should be chosen to give the required accuracy on the smallest cell frequency required. In practice, the situation is more complex because several cell frequencies are being simultaneously estimated and excesses in one cell over expectation must be balanced by deficits in other cells.

Given a sample distribution, it is natural to test a hypothesis that this sample has arisen from some hypothetical distribution. In Student's case, for example, his theoretical distribution, given by

$$f(t) = \frac{\Gamma\left(\frac{\upsilon+1}{2}\right)}{\sqrt{\upsilon}\sqrt{\pi}\Gamma\left(\frac{\upsilon}{2}\right)}\left(1+\frac{t^2}{\upsilon}\right)^{\upsilon+1/2},$$

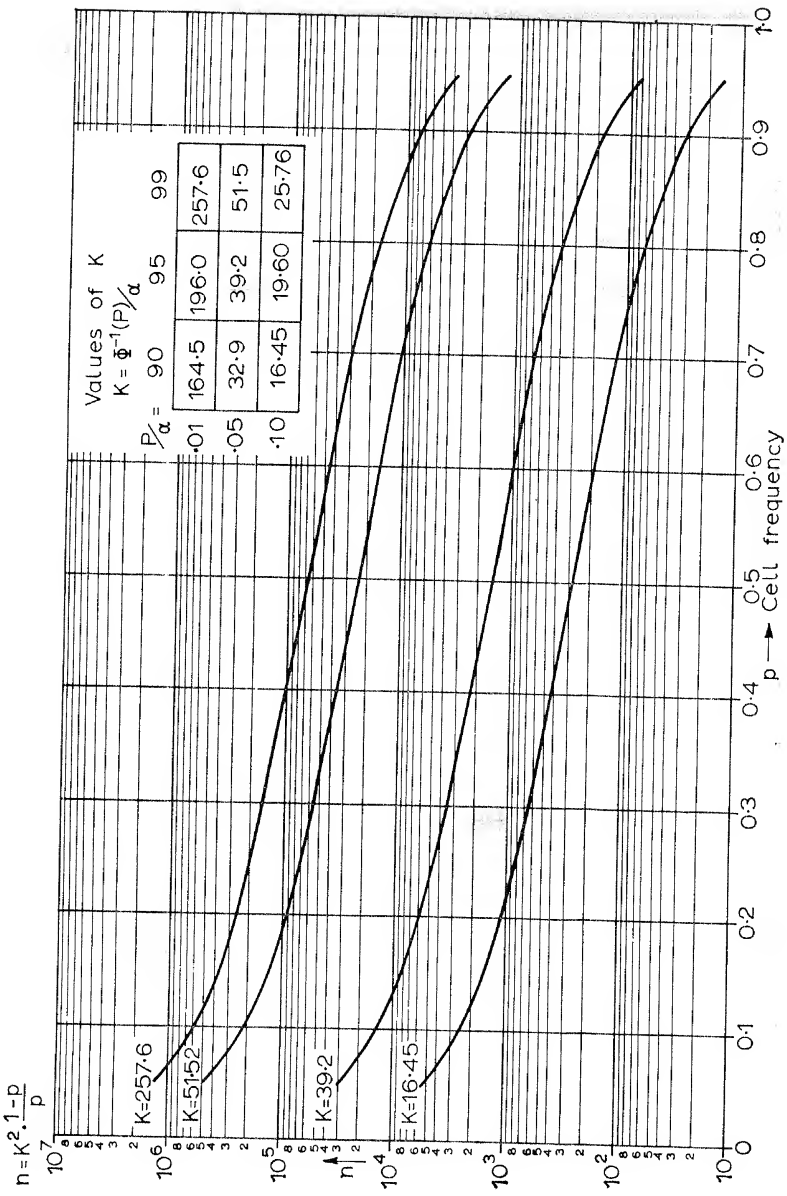

$n = K^2 \cdot \frac{1-p}{p}$
10^7
10^6
10^5
10^4
10^3
10^2
10^1
n
K=257·6
K=51·52
K=39·2
K=16·45
Values of K
$K = \Phi^{-1}(P)/\alpha$
P/α = 90 95 99
·01 164·5 196·0 257·6
·05 32·9 39·2 51·5
·10 16·45 19·60 25·76
0 0·1 0·2 0·3 0·4 0·5 0·6 0·7 0·8 0·9 1·0
p → Cell frequency

Figure 13

where υ is the degrees of freedom, would be the natural hypothesis. From this distribution the cell frequencies are calculated by quadrature and a comparison made by a χ^2 test.

7.1 Bivariate Distributions

In a similar way bivariate and even multivariate distributions may be sampled. No new principle is involved here but two or more statistics for each sample are calculated and the bivariate histogram developed.

In the same way, the sample size required for any given accuracy is determined by a similar calculation. However, the marginal distribution of each of the component statistics will also be of interest and must be reasonably well defined. It thus follows that the individual cell frequencies of the bivariate distribution are very much smaller than that for the corresponding marginal distribution. This implies that the total sample size required is very much larger in this case. Roughly the total number of cells required in the bivariate case is the square of that for the univariate case. In practice the resulting distributions obtained from such bivariate sampling experiments are too crude to give significant conclusions.

However, bivariate distributions are of great value in the more efficient estimation of irregular distributions.

7.2 Improving the Accuracy of Estimation of Sampling Distributions

In many of the situations, there are alternative statistics that may be used to estimate the parameter and in investigating by sampling experiments the distribution of one of these, it is a simple matter to calculate the other statistic at the same time. If both of these are useful estimators, we are likely, in successive samples, to obtain similar values from these estimates, i.e. these estimates are correlated. It is thus possible with very little extra work to obtain a bivariate distribution of two correlated statistics.

Now suppose that the marginal distribution of one of these statistics is known theoretically, then there must be some evidence concealed in departures of the actual sample distribution from the theoretical, and it should be possible to correct the sample values to take account of the deviations and thus improve the accuracy of the estimation of the sampling distribution of the first statistic. This technique was developed by Fieller and Hartley, and is illustrated in their paper by the joint distribution of the range and the sample standard deviation from independent samples from the normal distribution.

We sub-divide the range for the two statistics x and y, say, into intervals and denote the number in a sample falling in the ith y-category and the jth x-category by n_{ij}. The probability associated with this cell is denoted by p_{ij}. The marginal frequencies are given by

$$n_{\cdot j} = \sum_i n_{ij} \qquad p_{\cdot j} = \sum_i p_{ij}$$

$$n_{i\cdot} = \sum_j n_{ij} \qquad p_{i\cdot} = \sum_j p_{ij}$$

$$n = \sum_i n_{i\cdot} = \sum_j n_{\cdot j}$$

Also

$$\sum_i p_{i\cdot} = \sum_j p_{\cdot j} = 1$$

The conditional probability of a sample falling in the cell (i, j) given that it falls in the column j is $p_{ij}/p_{\cdot j}$, and is estimated by $n_{ij}/n_{\cdot j}$.

Now

$$p_{ij} = p_{\cdot j}\frac{p_{ij}}{p_{\cdot j}} \quad \text{and} \quad p_{i\cdot} = \sum_j p_{ij} = \Sigma p_{\cdot j}\frac{p_{ij}}{p_{\cdot j}}$$

and so an estimate of $p_{i\cdot}$ is

$$\hat{p}_i = \sum_j p_{\cdot j}\frac{n_{ij}}{n_{\cdot j}}$$

when $p_{\cdot j}$ is known.

Unfortunately if $n_{\cdot j}$ is zero, so must n_{ij} be, and the ratio is ill defined. In this case the estimate

$$\hat{p}_{ij} = p_{\cdot j}\frac{n_{i\cdot}}{n}$$

is used.

The complete estimating procedure can be summed up in the equation

$$\hat{p}_i = \Sigma p_{\cdot j} u_{ij}$$

where

$$u_{ij} = \frac{n_{ij}}{n_{\cdot j}} \qquad n_{\cdot j} > 0$$

$$= \frac{n_{i\cdot}}{n} \qquad n_{\cdot j} = 0$$

This formula leads to a slight bias in $\hat{p}_i$ since the expected value of u_{ij} is biased. In fact, it can be shown that

$$E(u_{ij}) = p_{ij}/p_{\cdot j} + (1 - p_{\cdot j})^{n-1}(p_{i\cdot} - p_{ij}/p_{\cdot j})$$

Values of Median	*Values of Mean (x)*															*Total Frequencies for y*		
(*y*)	6	7	8	9	10	11	12	13	14	15	16	17	18	19	*Observed*	$N\hat{p}_t$	*Expected*
5		1													1	1·9	1·9
6		2	1												3	4·7	6·2
7	2	1	8	6	3	1									21	21·6	16·8
8		1	11	11	15	4									42	41·0	38·3
9			9	20	21	10	1								61	59·9	71·5
10				10	54	34	14	1							113	109·3	110·4
11			1	10	30	48	43	7							139	140·1	139·4
12			2	3	11	34	56	26	8	1					141	142·8	144·1
13					2	8	38	58	22	2	1				131	131·6	121·9
14					1	3	14	28	29	9	5				89	89·5	84·5
15						1	2	8	18	8	3				40	40·8	48·1
16								3	3	8	5	1			20	19·9	22·6
17									2	4	4	3			13	12·0	8·8
18								1		1	1				3	2·9	2·9
19												1		1	2	1·0	0·8
Observed	2	5	32	60	137	143	168	132	82	33	19	5	0	1	819		
TOTALS Expected	2·4	9·5	28·5	65·6	116·5	158·1	165·6	132·8	81·7	38·7	14·1	3·9	0·8	0·2			

Hence

$$E(\hat{p}_{i\cdot}) = p_{i\cdot} - \sum_j (1-p_{\cdot j})^{n-1}(p_{ij} - p_{i\cdot}p_{\cdot j})$$

The bias

$$\sum_j (1-p_{\cdot j})^{n-1}(p_{ij} - p_{i\cdot}p_{\cdot j})$$

$$\sim \sum_j \frac{(p_{ij} - p_{i\cdot}p_{\cdot j})}{1-p_{\cdot j}} e^{-np_{\cdot j}} \sim \sum_j e^{-np_{\cdot j}}$$

This can be made negligible if the sample size n and cell boundaries are chosen so that $np_{\cdot j} \geqq 8$ (say). This can be achieved in advance of sampling since $p_{\cdot j}$ are known.

The advantage of this formula is, of course, a reduced variance of $\hat{p}_{i\cdot}$. This, by some tedious algebra, can be shown to be given by

$$V(\hat{p}_{i\cdot}) \simeq p_{i\cdot}(1-p_{i\cdot})/n - \sum_j (p_{ij} - p_{i\cdot}p_{\cdot j})^2/np_{\cdot}$$

$$+ \sum_j p_{ij}(1-p_{\cdot j})(1-p_{ij}/p_{\cdot j})/n^2 p_{\cdot j} + \sum_j (1-p_{\cdot j})^{n-2}(p_{ij} - p_{i\cdot}p_{\cdot j})^2$$

The last term of the expression may be ignored as it is of the same order of magnitude as the bias. The third term can be ignored relative to the second for large n. The first item is the ordinary binomial variance and the second represents the main variance reduction.

If x and y are independent then, by definition, $p_{ij} = p_{i\cdot}p_{\cdot j}$ and the variance reduction is identically zero. At the other extreme of complete dependence where $p_{ij} = \delta_{ij}p_{\cdot j} = \delta_{ij}p_{i\cdot}$ then the estimate $\hat{p}_{i\cdot}$ has zero variance. In the intermediate cases, if x and y are highly correlated, there will be a considerable reduction in the variance of $\hat{p}_{i\cdot}$.

We illustrate this procedure with the estimation of the distribution of a median of a sample of three items from a normal population, using the known distribution of the mean as a control variable.

The table on page 92 gives the result for 819 samples of 3 and tabulates the crude and adjusted cell frequencies and the true values taken from tables.

CHAPTER 8

FLOW DIAGRAMS

The description of sampling procedures in ordinary English is rather word-consuming, and it is desirable to develop an abbreviated language for these descriptions. Further, since for large-scale sampling experiments computers are almost a necessity, it is desirable to develop a language that is suitable for describing the problem for a programmer.

This is achieved by the flow diagram. There are several conventions used for these diagrams, which describe the successive steps of the calculation, and we shall adopt the system involving minimal conventions and as general as possible and not oriented towards any particular computer.

The success of the automatic computer hinges on the fact that all extensive calculations can be described as repetitions of cycles of comparatively short calculations. Thus, each order in the programme is used many times over. If the programme had to consist of a string of orders obeyed serially, the time to transcribe the orders might easily exceed that necessary to do the calculation by hand. Such cycles of orders must not be repeated indefinitely, and so the order code of all machines contains orders that allow tests to be made on the quantities produced by the calculation, and as a result of the tests, the order of execution of the orders may be determined.

Thus the programme consists of blocks of orders which effect the calculations required, each terminated by a test (or tests) which directs the computer to the appropriate next block.

We shall use a rectangular block to represent the calculating orders and will write a brief description of the calculation in the block. The test at the end of each block will be written in an oval and branches from the circle will indicate the possible blocks of programme next executed and the conditions necessary to follow each route.

Thus to print the 100 means, each from a sample of 5, from a normal distribution, the flow diagram would appear as in Figure 14 (a) on page 95.

Usually, but not universally, the choice of routes at each branch point is limited to two, and in this case the route followed when the test named is satisfied is marked with a 0 and the route followed on failure with a *X*.

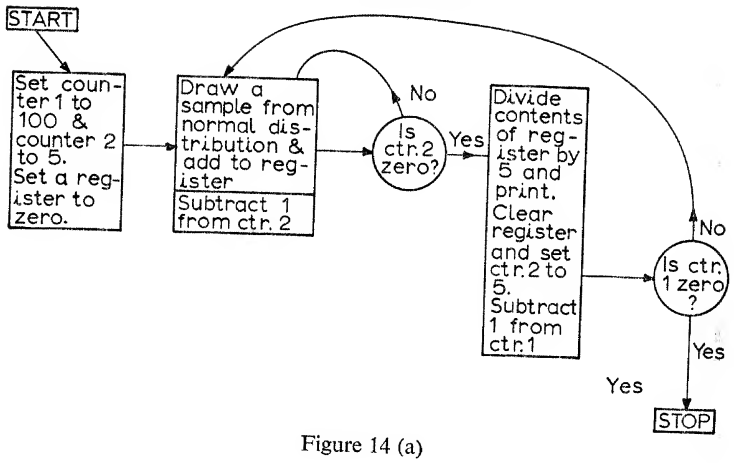


Figure 14 (a)

The description in the calculating block can be shortened when it is recognised that the numbers involved are stored in registers during and after the calculation. If these registers are given letters as names, the calculation can often be conveniently described by specifying the contents of the register at the conclusion of the calculation.

It is also advantageous to distinguish between variables and indices. The latter are typically counters and marks to help specify the course of the calculation. Indices are denoted by lower-case letters and variables by capitals. Similar quantities are denoted by the same letter and distinguished by numerical suffices. In this convention the previous flow diagram becomes:

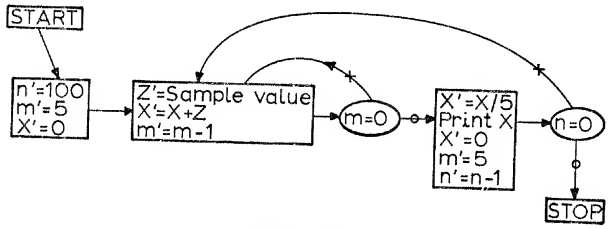


Figure 14 (b)

Alternatively we could set m and n initially to zero, count up and test for $m = 5$ and $n = 100$. The way displayed is more convenient as most computers can make the test for a zero value immediately. Whichever convention is adopted, the direction of count is obvious. Often the count

is *immediately* followed by a test. In these cases, the abbreviation '*ct m*' will be used for both the count and the test.

When the calculation is set out in this abbreviated notation, it is often easier to see possibilities of improvement. In the example, the pair of operations $X' = 0$ $m' = 5$ occur twice and this enables the procedure to be modified to:

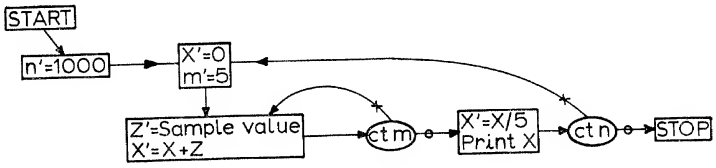


Figure 15

It is not always necessary to specify each process in full in one flow diagram. In the example of the last chapter used to illustrate the use of the mean as a control variable, the determination of the median in terms of elementary operations is quite difficult. However, the main flow diagram can dismiss this by a single statement.

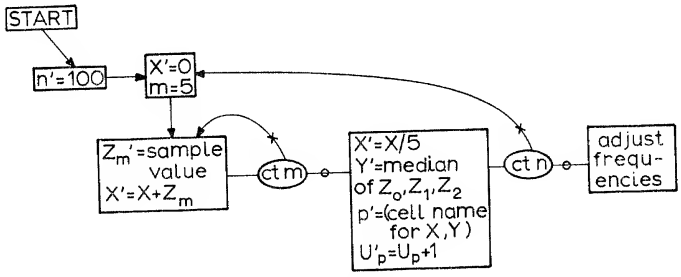


Figure 16

Several additional notational points are raised by this example. The suffix on the variable Z is an index adjusted by the programme. The frequencies being collected are indexed by p determined by the variables X, Y, which involves a further calculation.

The calculation 'Y' = median of Z_0, Z_1, Z_2' has its own sub-flow diagram as shown in Figure 17 on page 97.

The reader may find it of interest to examine the case when two or more Z's are equal.

For large sample sizes, this simple technique would involve an interminable number of different tests and a more systematic method is required.

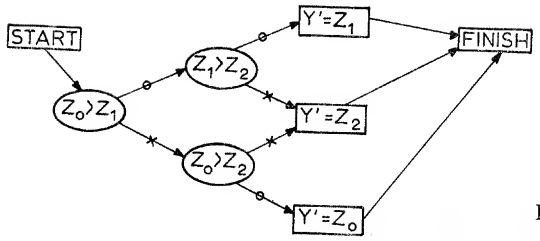


Figure 17

Consider the case of an odd sample size $2m+1$. Then there are m sample values greater than or equal to the median value. A method would be to try each value in turn until one satisfying this condition is found. This has two disadvantages: if the sample contains several equal medial values difficulties arise; information from former trials is wasted in later trials. The first disadvantage is avoided if both the number of sample values greater than the trial one, and the number less than it are counted. Both these counts must be less than or equal to m for the median.

Trials can be avoided if bounds on the median are adjusted after each trial and values outside the bounds abandoned.

The flow diagram to describe this is as follows:

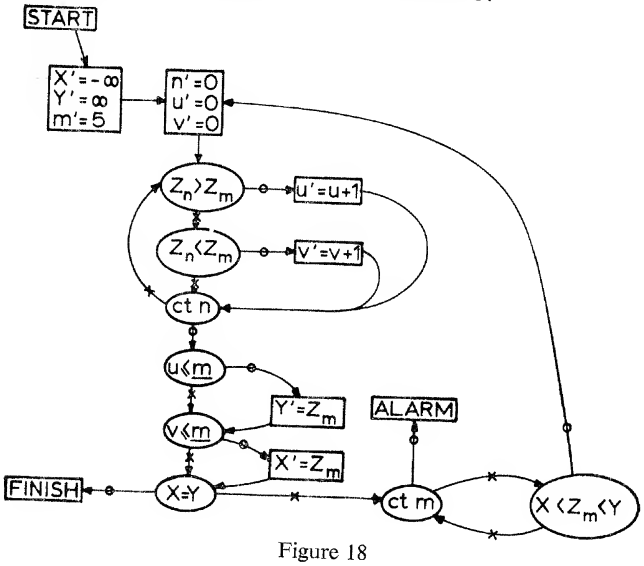


Figure 18

The value m is underlined, where mentioned in the diagram, to distinguish it from the register m or its contents.

The box 'ALARM' illustrates another feature and advantage of this approach. It makes it easier to incorporate tests for errors. In this case, the adjustment of the count on m is automatically followed by a test and its success clearly means an error, since it is now not possible to find the median. Without the test, the error would have continued to search over non-existent sample values to find the median.

To illustrate the working of this cyclic method, consider the sample of 9 values ($m = 4$)

m	0	1	2	3	4	5	6	7	8
Z_m	7	13	5	5	9	9	8	5	8
Rank	6	1	7	7	2	2	4	7	4

The values of m, Z_m, u, v, X, Y on each successive trial are tabulated below

	m	Z_m	u	v	X	Y
initially	0	—	0	0	$-\infty$	∞
	0	7	5	3	7	∞
	1	13	0	8	7	13
	4	9	1	6	7	9
	6	8	3	4	8	8

Similarly, a sub-flow diagram for the adjustment of the cell frequencies and the marginal frequencies is required. All computers have a linear ordering of their registers and the required two-dimensional ordering must be converted to a one-dimensional one. Suppose the array is $N \times M$, then the (i, j)th cell frequency is stored in the

$$p\text{th} = ((i-1)N+j)\text{th}$$

register of a group reserved for that purpose. Denote these registers by U_p. Denote the registers holding the marginal frequencies by V_i and W_j where i and j are the cell labels with upper values I and J and suppose the common cell widths are a and b respectively, then the cell counting is merely

$i' = X/a$ (unrounded division)
$i' = \min(i, I)$
$j' = Y/b$ (unrounded division)
$j' = \min(j, J)$
$p' = (i-1)N+j$
$U'_p = U_p+1$
$V'_i = V_i+1$
$W'_j = W_j+1$

Store the theoretical marginal frequencies for the mean in registers L_i. Then the flow diagram for the adjusting process will be:

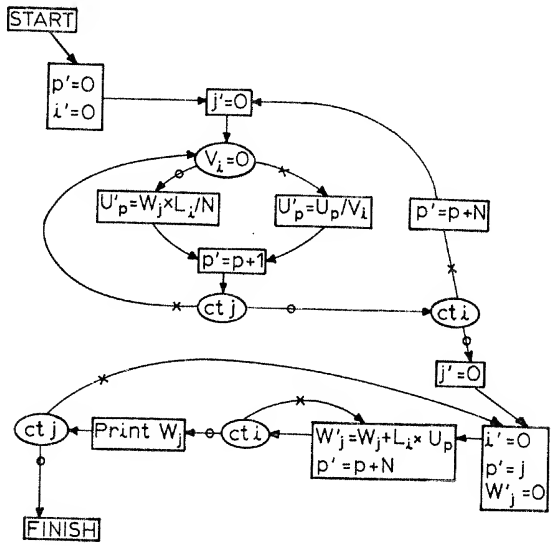


Figure 19

Although this flow diagram procedure may seem rather cumbersome, it is an essential step in preparing a computer programme and even if the calculation is to be done by hand, it ensures that the detail of the calculation is understood before a computing schedule is drawn up. For the automatic computer, every contingency must be considered, and this also saves time on manual calculation when an unusual situation arises.

These diagrams will be used extensively in later chapters, but generally the detailed sub-flow will be omitted.

The second stage of cementing these diagrams into computer programmes or of drawing up computing schedules will depend upon the equipment being used and will not be considered further.

CHAPTER 9

ESTIMATION

An earlier chapter has dealt with the problem of determining a distribution by a sampling process. Often, the aim of the experiment is less ambitious and only the mean (and perhaps the standard deviation) is required. A classic example of this arises in the use of range as a measure of dispersion. The range is clearly a monotonic increasing function of the sample size as increasing the sample size increases the probability of obtaining at least one large deviation from the mean; what is the functional relationship between the mean range and the sample size?

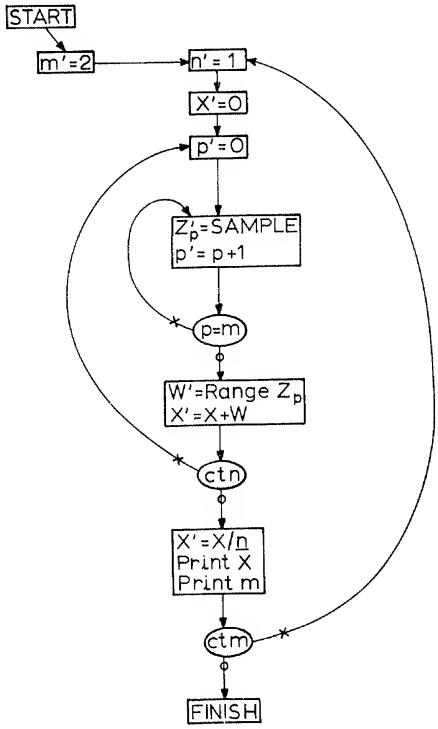


Figure 20

In fact, this relationship is known analytically, but several object lessons can be learned from an attempt to evaluate this relationship by sampling. The first crude method would simply consist of drawing many samples of size n, evaluating the range for each, and use their average as an estimate for the graph of expected range against sample size. This would be repeated for various n. A flow diagram of this process would be as shown in Figure 20 on page 100.

The calculation of the range follows a similar scheme to that proposed for the calculation of the median.

It now becomes clear that this process can be combined with the sampling process and thus eliminate the storage of the sample values. Maximum and minimum values are adjusted after each sample. This raises the possibility of using the partially complete sample for one value of n to give an estimate of range from a smaller sample. Such estimates will be positively correlated with each other, but this will ensure that differences between mean ranges for different sample sizes are estimated with greater precision. The amount of sampling required is reduced to that for the largest sample size to be investigated.

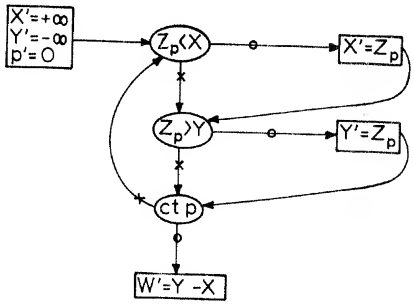


Figure 21

A little care is needed with the initial conditions since a sample of one cannot give a range. An outline of the calculation is given in Figure 22 on p. 102.

We now turn to improvements in the sampling procedure. Suppose the mean of the distribution being sampled is known to be the value μ. If the mean of the sample chosen is greater than zero, we will expect a larger maximum than in the contrary case. If we were concerned with estimating the maximum of the sample, then a more sensitive statistic to use is

$$U = \max_p Z_p - \Sigma Z_p/n \qquad E(U) = E(\max_p Z_p) - \mu$$

Our estimate is then $U+\mu$. U has smaller sampling variance than $\max_p Z_p$ and so greater precision is obtained for a given sample size.

Similarly $V+\mu$ is a better estimate of $\min_p Z_p$ where

$$V = \min_p Z_p - \Sigma Z_p/n$$

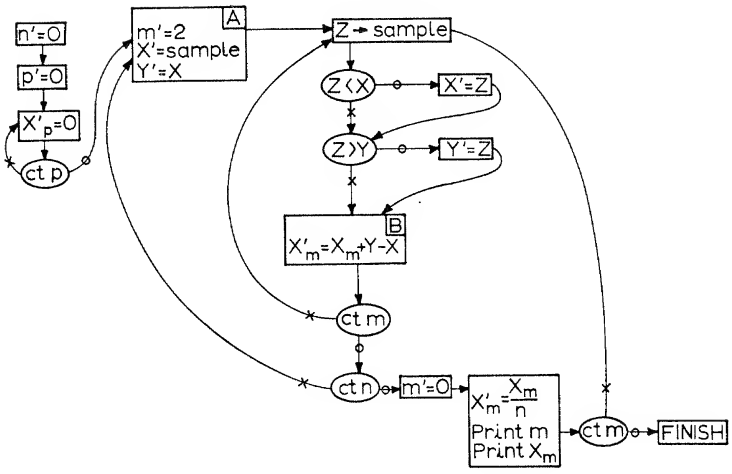


Figure 22

Unfortunately, the estimate of the range

$$U - V \equiv \max_p Z_p - \min_p Z_p = \text{range } Z_p = W \quad \text{say}$$

eliminates the mean.

However, we can use this *control variate* method with some other measure of the spread. If the standard deviation of the distribution is known to be σ we can use the statistic

$$W_p - \sqrt{\frac{\Sigma(Z_p - \bar{Z})^2}{n-1}} + \sigma \tag{1}$$

as an estimate of the range from each sample where $\bar{Z} = \Sigma Z_p/n$. The adjustment to the flow diagram is

(i) add two instructions $U' = 0$, $V' = 0$ to box A

(ii) replace box B by

$$U' = U + Z$$
$$V' = V + Z^2$$
$$W' = \sqrt{\frac{mV' - U^2}{m(m-1)}}$$
$$X'_m = X_m + Y - X - W$$

We may improve on this method by noticing that

$$W - k\sqrt{\frac{\Sigma(Z_p - \bar{Z})^2}{n-1}} + k\sigma \tag{2}$$

is a valid estimate of the range for any k and thus k can be chosen to reduce the sampling variance. We require minimal sampling variance and so a good choice of k is that which makes the *estimated* sampling variance minimum, i.e. we choose k so that

$$\Sigma(W - kS - \overline{W - kS})^2$$

is minimised. We can use the method of least squares to estimate k. This technique will be discussed later on in the chapter.

It is important to note that any error in this estimate from the theoretical optimal will not affect the validity of our estimate (1) but merely decreases its efficiency.

The changes to the flow diagram to use this procedure are left to the reader.

If the distribution being sampled is symmetric, we may make use of the sample mean in another way. We have

$$E(W) = E(W|\bar{Z} > \mu).Pr(\bar{Z} > \mu) + E(W|\bar{Z} \leqq \mu).Pr(\bar{Z} \leqq \mu)$$

If, as usual, there is no lump probability at $Z = \mu$ the symmetry of the distribution of Z implies that of $\bar{Z}$ so that

$$Pr(\bar{Z} > \mu) = Pr(\bar{Z} \leqq \mu) = \tfrac{1}{2}$$

The procedure is amended to measure $\bar{Z}$ for each sample and to stratify the data into the two classes. If the number of samples for which $\bar{Z} < \mu$ is n_1, and the number for which $\bar{Z} \geqq \mu$ is n_2 $(n = n_1 + n_2)$ the best estimate to use is

$$\frac{1}{2}\left\{\frac{\sum_{\bar{Z}<\mu} W}{n_1} + \frac{\sum_{\bar{Z}\geqq\mu} W}{n_2}\right\}$$

These two variance-reducing methods can be combined by using (3) with the statistics (1) or (2).

If the range is found, as proposed, from the difference of $X = \min_p Z_p$ and $Y = \max_p Z_p$, a more effective stratification is possible.

$$E(W) = E(W|X > \mu,\ Y > \mu).Pr(X > \mu,\ Y > \mu)$$
$$+E(W|X \leqq \mu,\ Y > \mu).Pr(X \leqq \mu,\ Y > \mu)$$
$$+E(W|X \leqq \mu,\ Y \leqq \mu).Pr(X \leqq \mu,\ Y \leqq \mu)$$

Now if $X > \mu$ and $Y > \mu$ then $Z_p > \mu$ for all p and the probability of this is $(\frac{1}{2})^m$ for sample size m.

Similarly,

$$Pr(X \leqq \mu,\ Y \leqq \mu) = Pr(X > \mu,\ Y > \mu) = (\tfrac{1}{2})^m$$

and hence $$Pr(X \leqq \mu,\ Y > \mu) = 1-(\tfrac{1}{2})^{m-1}$$

If the number of samples in the classes is given by:

$$X > \mu,\ Y > \mu \qquad n_1$$
$$X \leqq \mu,\ Y > \mu \qquad n_2$$
$$X \leqq \mu,\ Y \leqq \mu \qquad n_3$$

then the best estimate is

$$(\tfrac{1}{2})^m\left\{\frac{\sum_{X>\mu,Y>\mu} W_p}{n_1}+\frac{\sum_{X\leqq\mu,Y\leqq\mu} W_p}{n_3}\right\}+\{1-(\tfrac{1}{2})^{m-1}\}\left\{\frac{\sum_{X\leqq\mu,Y>\mu} W_p}{n_2}\right\}$$

The advantage of this stratification procedure is that the chance variations of the sample frequencies from their theoretical values are eliminated and this reduces the total variability of the resulting estimate.

The two stratifications can be combined. It is clear that

$$Pr(X > \mu,\ Y > \mu,\ \bar{Z} \leqq \mu) = Pr(X \leqq \mu,\ Y \leqq \mu,\ \bar{Z} > \mu) = 0$$

and by symmetry

$$Pr(X \leqq \mu,\ Y > \mu,\ \bar{Z} \leqq \mu) = Pr(X \leqq \mu,\ Y > \mu,\ \bar{Z} > \mu) = \tfrac{1}{2}-(\tfrac{1}{2})^m$$

Then the estimate used is

$$(\tfrac{1}{2})^m\left\{\frac{\sum_{X>\mu,Y>\mu} W_p}{n_1}+\frac{\sum_{X\leqq\mu,Y\leqq\mu} W_p}{n_2}\right\}$$
$$+\{\tfrac{1}{2}-(\tfrac{1}{2})^m\}\left\{\frac{\sum_{X\leqq\mu,Y>\mu,Z<\mu} W_p}{n_3}+\frac{\sum_{X\leqq\mu,Y>\mu,Z\geqq\mu} W_p}{n_4}\right\}$$

If the sampling distribution of an auxiliary correlated statistic is known, then stratification with respect to this statistic can be used. For

example, the distribution of s^2 is known for a normal distribution so that if the expected value of range of a sample from a normal distribution is under investigation, we may use the formula

$$E(W) = \sum_j E(W|S_j)P(S_j)$$

where S_j is the jth division of the range $(0, \infty)$ and $P(S_j)$ is the probability that s^2 falls into S_j. $E(W|S_j)$, the expected value of W subject to the sample s^2 falling in S_j is estimated by the average of the sample W's whose associated s^2 do fall in S_j.

In large samples, the numbers falling into the different strata will be approximately proportional to the probabilities associated with them. The advantage of the stratification is that the chance variability of these numbers no longer affects the variability of the estimate.

However, it is still possible, particularly if a fine stratification is used, that a stratum may fail to contain a sample at all. In this case, the estimate of conditional expected values is no longer available.

In the analogous case, when estimating a frequency distribution using a control variable, an alternative estimate was used and this led to bias. In the case under discussion, a similar expedient can be used substituting for the indeterminate values the unconditional average.

However, this will also bias the estimate, and an alternative procedure is desirable. The simplest expedient is to sample until every stratum has a representative (and the minimum sample size has been reached). This leads to variable sample size and the exact statistical analysis of the accuracy of estimate is considerably complicated by this fact. In practice, for large samples, the probability of a sample size beyond the minimum is low and the effect of the variability of sample size is of no importance.

However, it may be more economical in time and energy to fix the sample size and modify the sampling procedure so that the sample size in each stratum is predetermined.

The sampling techniques illustrated in this example can be generalised and subjected to a certain amount of mathematical analysis.

We denote the sample values $x_1, x_2, \ldots, x_m$ by the vector $\boldsymbol{x}$ whose p.d.f. is denoted by $p(\boldsymbol{x})$. Usually this is of the form

$$p(\boldsymbol{x}) = \prod_{i=1}^{m} p(x_i)$$

but this is not necessary to the arguments that follow:

The statistic is a function of x denoted by $Z(x)$. Again usually this is a symmetric function of $x_1, x_2, \ldots, x_m$ but this fact is not utilised in the main part of the theory.

The quantity to be estimated is the expected value of Z, i.e.

$$\bar{Z} = \int_R Z(x)p(x)\,dx$$

where R is the total region for which $p(x)$ is non zero. Usually this is a m-dimensional Euclidean space but can, for what follows, be any region.

The stratification is represented by dividing R into k sub regions $R_j (j = 1, 2, \ldots, k)$.

We define

$$p_j = \int_{R_j} p(x)\,dx \quad \text{the probability of a sample falling in } R_j$$

$$\left.\begin{aligned} P_j(x) &= p(x)/p_j \qquad x \varepsilon R_j \\ &= 0 \qquad\quad \text{otherwise} \end{aligned}\right\} \text{the conditional p.d.f. of } x \text{ given that it is in } R_j$$

$$\left.\begin{aligned} \bar{Z}_j &= \int_R Z(x)p_j(x)\,dx \\ &= \frac{1}{p_j}\int_{R_j} Z(z)p(x)\,dx \end{aligned}\right\} \text{the conditional expected value of } Z \text{ given that } x \text{ is in } R_j$$

Given n samples $x_1, x_2, \ldots, x_n$, the corresponding values of Z are denoted by $Z_1, Z_2, \ldots, Z_n$. For a stratified sample containing n_i items in R_i the values are written as Z_{ij} $(j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k)$.

The simple estimate of $\bar{Z}$ is

$$\hat{Z}_1 = \frac{1}{n}\Sigma Z_i$$

Now $E(Z_i) = \bar{Z}$ and its sampling variance is

$$\sigma^2 = \overline{(Z-\bar{Z})^2} = \int_R (Z-\bar{Z})^2 p(x)\,dx = \int_R Z^2 p(x)\,dx - \bar{Z}^2$$

The variance of $\bar{Z}_j$ is similarly defined as

$$\sigma_j^2 = \overline{(Z_j-\bar{Z}_j)^2} = \int_R Z^2 p_j(x)\,dx - Z_j^2$$

$$= \frac{1}{p_j}\int_{Rj} Z^2 p(x)\,dx - \bar{Z}_j^2$$

Now $$\sigma^2 = \int_R (Z-\bar{Z}_j+\bar{Z}_j-\bar{Z})^2 p(x)\,dx$$

$$= \int_R (Z-\bar{Z}_j)^2 p(x)\,dx + \int_R (\bar{Z}_j-\bar{Z})^2 p(x)\,dx$$

$$= \sum_j p_j\{\sigma_j^2+(\bar{Z}_j-\bar{Z})^2\}$$

Thus $$V(\hat{Z}_1) = \frac{\sigma^2}{n} = \frac{1}{n}\left\{\sum_j p_j\sigma_j^2 + \sum_j p_j(\bar{Z}_j-\bar{Z})^2\right\}$$

The estimate from a stratified sample is

$$\hat{Z}_2 = \sum_i p_i\left(\frac{\sum_j Z_{ij}}{n_i}\right)$$

since $$E(Z_{ij}) = \bar{Z}_i$$

$$E(\hat{Z}_2) = \sum_i p_i\bar{Z}_i = \bar{Z}$$

$$V(\hat{Z}_2) = \sum_i \frac{p_i^2}{n_i^2}(n_i\sigma_i^2) = \sum_i \frac{p_i^2}{n_i}\sigma_i^2$$

On a freely chosen sample n_i will be a variable but in large samples $n_i \simeq np_i$ and

$$V(\hat{Z}_2) = \frac{1}{n}\sum_i p_i\sigma_i^2 < \frac{\sigma^2}{n} = V(\hat{Z}_1)$$

$$V(\hat{Z}_1) - V(\hat{Z}_2) = \frac{1}{n}\sum_i p_i(\bar{Z}_i-\bar{Z})^2$$

If the sample is restricted so that the number in each stratum is exactly np_i the above formula becomes exact. However, if restricted sampling is possible a better choice of the n_i is possible.

We require to minimise $\Sigma\frac{(p_i\sigma)^2}{n_i}$ subject to $\Sigma n_i = n$. The optimum values of n_i can be found by the method of Lagrangian multipliers as

$$n_i = \frac{p_i\sigma_i}{\sum_i p_i\sigma_i}.n$$

The variance of this estimate $\hat{Z}_3$ say is

$$V(\hat{Z}_3) = \frac{1}{n}\left(\sum_i p_i\sigma_i\right)\left(\sum_i \frac{(p_i\sigma_i)^2}{p_i\sigma_i}\right) = \frac{\bar{\sigma}^2}{n}$$

since

$$\sum_i p_i\sigma_i^2 = \sum_i p_i(\sigma_i-\bar{\sigma})^2+\bar{\sigma}^2$$

$$V(\hat{Z}_2)-V(\hat{Z}_3) = \frac{1}{n}\sum_i p_i(\sigma_i-\bar{\sigma})^2$$

Often a restricted sample can only be taken by a rejection process. In this case, of course, nothing can possibly be gained by throwing away sample values except the labour of calculating Z values for the samples. It is very rare for the sampling process and the test if it lies in any given region to be negligible work compared to the work required to calculate Z.

However, if a scheme of restricted sampling not involving rejection can be devised then the optimum stratification of the sample is rewarding. The exact value of stratification cannot readily be stated since nothing can be said about the relative magnitude of the components of the partition

$$\sigma^2 = \Sigma p_i(\bar{Z}_i-\bar{Z})^2+\Sigma p_i(\sigma_i^2-\bar{\sigma})^2+\bar{\sigma}^2$$

The possibility exists in theory but does not often materialise in practice that a stratification can be devised from which the $\bar{Z}_j$ can be calculated, but the p_j are unknown.

The sampling procedure then consists of determining for each sample the value of j, $j(i)$, say, and using the estimate

$$V(\hat{Z}_4) = \frac{1}{n}\sum \bar{Z}_{j(i)}$$

Now

$$V(\bar{Z}_{j(i)}) = \sum_j p_j\bar{Z}_j^2-\bar{Z}^2$$

$$V(\hat{Z}_4) = \frac{1}{n}\sum_j p_j(\bar{Z}_j-\bar{Z})^2$$

Another possibility is that both p_j and $\bar{Z}_j$ are known for some of the strata (if known for all the estimation problem is solved). Then restricted sampling in the remaining strata should be used and the quantity $\sum_U p_i\bar{Z}_j$ estimated and the final estimate obtained by adding $\sum_K p_i\bar{Z}_j$. (K and U

standing for the set of strata in which p_j and $\bar{Z}_j$ are known and unknown respectively.)

Other mixed situations exist and the modification required for any individual situation is easily made.

A final possibility is to make a random choice whether a sampled point will be used or not. Suppose the labour of calculating Z is large compared to the labour of selecting the sample $\boldsymbol{x}$. Then for each selected point the value of j can be determined and the point accepted with probability q_i and rejected with probability $1-q_j(\neq p_j)$. If accepted, $\bar{Z}$ is calculated, but otherwise not.

Put
$$Z = \bar{Z}/q_j \quad \text{if } \boldsymbol{x} \text{ accepted.}$$
$$= 0 \quad \text{if not.}$$

Then
$$E(Z) = \sum_j p_j\left(\frac{\bar{Z}_j}{q_j}q_j + 0(1-q_j)\right) = \sum_j p_j\bar{Z}_j = \bar{Z}$$

Use an estimate
$$\hat{Z}_5 = \frac{1}{n}\sum_i Z_i$$

Now
$$V(Z_5) = \sum_j p_j\left(\frac{Z_j^2}{q_j^2}q_j\right) - \bar{Z}^2 = \sum_j p_j\frac{\overline{Z_j^2}}{q_j} - \bar{Z}^2$$

If we choose q_j to minimise this quantity we obviously take $q_j = 1$ but this fails to reflect the reduction of calculation achieved by merely sampling and not calculating Z. Suppose that taking the sample and determining the strata is $100\phi\%$ of the total work. The expected amount of work per sample is

$$\sum_j p_j\{q_j + \phi(1-q_j)\} = \sum_j p_j(1-\phi)q_j + \phi$$

The optimum solution for q_j is

$$q = \min\,(1, k\sqrt{\overline{z_j^2}})$$

where k is chosen so that the expected amount of work per sample takes the required value.

If the strata probabilities are known and a restricted sampling procedure is possible, then the process is best modified.

(i) Pick a strata with probability p_j.

(ii) Choose to sample this strata with probability q_j.

This reduces the work still further for the samples which will not be used to provide a value of Z.

If this splitting process is used, it is best to divide the strata into two classes. The first class is sampled by an ordinary stratified procedure and the remainder treated by the splitting technique.

The general motivation for the splitting technique is that the strata so treated have very low σ_j's and a few samples from a stratum of this kind will estimate $\bar{Z}_j$ adequately, and the work of restricting a sample to the region is high and so the return on work done is low.

The splitting technique is often referred to as Russian roulette.

As an example, consider the evaluation of $E(e^{-R^4})$ where $R^2 = x^2+y^2$ and (x, y) are the co-ordinates of a random point in a square of side 2 with centre the origin. Take $k = 2$ and define

$$R_1 \text{ as } x^2+y^2 \leqq 1$$
$$R_2 \text{ as } x^2+y^2 > 1 \qquad -1 \leqq y \leqq 1$$
$$p_1 = \frac{\pi}{4} \qquad p_2 = 1-\frac{\pi}{4}$$

By symmetry we may restrict the sampling to the first quadrant.

Using stratified unrestricted sampling the flow diagram reads

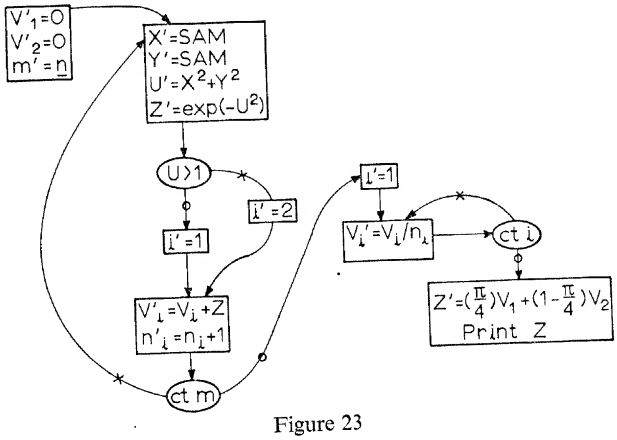


Figure 23

For restricted sampling from $R1$ we take x with p.d.f.

$$\frac{4}{\pi}\sqrt{1-x^2} \qquad \text{(distribution } R1\text{)}$$

y uniform in $(0, \sqrt{1-x^2})$ (distribution $S1$)

From $R2$ we take x with p.d.f.

$$(1-\sqrt{1-x^2})\Big/\left(1-\frac{\pi}{4}\right) \qquad \text{(distribution } R2\text{)}$$

and y uniformly in $(\sqrt{1-x^2}, 1)$ (distribution $S2$)

One scheme takes $n_1 \simeq \left[\frac{\pi}{4}n\right]$ $\quad n_2 \simeq \left[\left(1-\frac{\pi}{4}\right)n\right]+1$

The one is added to n_2 to ensure a sample size n, since $\frac{\pi}{4}n$ can never be an integer.

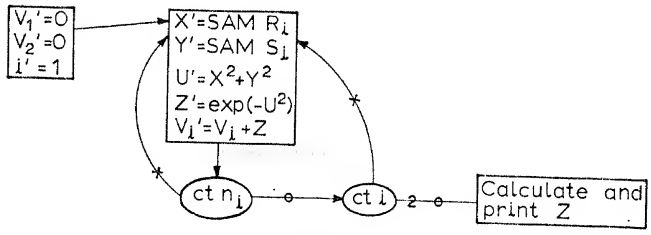


Figure 24

The optimum sample division for a restricted sampling scheme is difficult to determine since the evaluation of σ_1 and σ_2 is a more difficult problem than that of determining $E(Z) = E(e^{-R^4})$. Rough bounds can be found

$$\text{In } R_1 \qquad 1 \geqq Z \geqq e^{-1}$$
$$\text{In } R_2 \qquad e^{-1} \geqq Z \geqq e^{-4}$$

The ranges are the ratio $1-e^{-1}/(e^{-1}-e^{-4}) \simeq e-1$ and if we assume the standard deviations in the same ratio, we take

$$n_1 = \left[\frac{\frac{\pi}{4}(e-1)}{1+\frac{\pi}{4}(e-2)}n\right] \qquad n_2 = \left[\frac{1-\frac{\pi}{4}}{1+\frac{\pi}{4}(e-2)}n\right]+1$$

To apply Russian roulette note that

$$\text{In } R_1 \qquad 1 \geqq \overline{Z^2} \geqq e^{-2}$$
$$\text{In } R_2 \qquad e^{-2} \geqq \overline{Z^2} \geqq e^{-16}$$

Thus the splitting is done on region R_2. This is also advantageous since the sampling from R_2 is slightly more difficult. The estimation of ϕ is

rather difficult. Assume a value 0·1 and a 10% reduction in sampling. Then substituting in the general formula gives

$$q_2 = \frac{3{\cdot}2 - 0{\cdot}9\pi}{3{\cdot}6 - 0{\cdot}9\pi}$$

The practical difficulties of estimation of quantities such as $\overline{Z_j^2}$, σ_j, ϕ, etc., make the application of restricted and split samples of doubtful utility except in cases where a stratification giving some order of magnitude differences in σ_j can be easily made.

The evaluation of the relevant parameters required to design an efficient sampling procedure raises an estimation problem. In principle, the procedure should proceed as a two-stage sampling method in which the primary purpose of the first stage is to enable a second stage to be planned so that the overall sample will have been selected on an optimum partition between the strata.

Unfortunately the variability of an estimate of a variance is usually higher than that of a mean and so the secondary parameters can only be estimated from a sub-sample with very poor precision. If the appropriate optimum allocation of sampling between the strata involves very low sampling rates for some strata the original sub-sample may contain more than required in order to estimate the secondary parameters.

The ideal procedure would be to make provisional estimates of both primary and secondary parameters as each sample (or batch of samples) is taken and to evolve the design in the light of the best estimates available.

This discussion has assumed the amount of work (represented by the size of sample taken) that is available is fixed and has sought to utilise this to minimise the variability of the estimate used. This is artificial and it might be more appropriate in many cases to fix the variance required and design the sampling scheme to minimise the sample size required. Formally these two problems are identical, but in the latter the absolute values of the secondary parameters are needed, whereas in the former relative values will often be sufficient.

Thus an evolutionary design should terminate when a required accuracy has been achieved. Sequential estimating schemes of this kind using stratification have received very little attention in the literature, but the simple sequential unstratified sampling scheme has received considerable attention.

Since any sequential scheme has a stopping rule which makes the sample size variable, the actual standard deviation of the resulting

estimate will be variable. A different definition of accuracy is required and a confidence interval of specified length and confidence is employed.

Suppose that the mean of a normal distribution $N(\mu, \sigma)$ is to be found by sampling. The mean of n-observations will have a normal distribution $N(\mu, \sigma/\sqrt{n})$. Define t_α by

$$\frac{1}{\sqrt{2n}} \int_{-t_\alpha}^{t_\alpha} e^{-\frac{1}{2}x^2}\, dx = 1-\alpha$$

then the probability if the estimate of the mean is $\bar{x}$, that the interval

$$(\bar{x}-t_\alpha\sigma/\sqrt{n} \quad , \quad \bar{x}+t_\alpha\sigma/\sqrt{n})$$

does not contain the true mean μ is α. If we require the interval to be of length l this gives the equation for n

$$n = \left(\frac{2t_\alpha\sigma}{l}\right)^2$$

If this is not an integer the next highest integer should be used.

This assumes that σ is known. Otherwise this must be estimated from the data.

If a fixed sample is taken the appropriate confidence interval is found by using Student's distribution. Let t_α now be the critical *t-value* for enclosing a probability of $1-\alpha$. Calculate

$$t = \frac{(\bar{x}-\mu)\sqrt{n}}{s}$$

This gives a confidence interval

$$\left(\bar{x}-\frac{t_\alpha s}{\sqrt{n}} \qquad \bar{x}+\frac{t_\alpha s}{\sqrt{n}}\right)$$

The analogous sequential procedure would at first sight take a rule to stop sampling when

$$n \geqq \left(\frac{2t_\alpha^{(n-1)}S_n}{l}\right)^2$$

where S_n is the estimated standard deviation on n observations and $t_\alpha^{(n)}$ the critical t-value for Student's test on n degrees of freedom. However,

this is not theoretically correct as the probability of stopping involves the successive events

$$S_2^2 \geqq 2(\tfrac{1}{2}l)^2/t_\alpha^{(1)}$$
$$S_3^2 \geqq 3(\tfrac{1}{2}l)^2/t_\alpha^{(2)}$$
$$\vdots$$
$$S_n^2 \leqq n(\tfrac{1}{2}l)^2/t_\alpha^{(n)}$$

Since the statistics $S_1, S_2, \ldots, S_n$ are not independent, the probability of this joint event is not easily calculated and is certainly not $1-\alpha$ as supposed by the confidence interval argument. However, for large values of n, Anscombe and others have shown that the effect of the sequential nature of the process can be accommodated to a sufficient accuracy for practical application by a simple adjustment.

The recalculation of S_r^2 after taking each sample is troublesome and can be avoided by rewriting the stopping rule as a restriction on the sum of squares about the mean.

This can be done most conveniently by means of the Helmholtz transformation.

Suppose the sample is $x_1, x_2, \ldots, x_n$

Then
$$\sum_{i=1}^{n}(x_1-\bar{x})^2 = \sum_{i=1}^{n-1} u_i$$

where
$$u_i = \frac{1}{i(i+1)}\left\{ix_{i+1} - \sum_{j=1}^{i} x_i\right\}^2 \qquad i = 1, 2, \ldots, n-1$$

Thus each additional sample taken enables a new u_i to be calculated and does not change the values of the previous u-values.

The modified stopping rule in terms of these u-variables is found to be

$$\sum_{i=1}^{n-1} u_i \leqq \frac{l^2}{4t_\alpha^2} n(n-2{\cdot}676-\tfrac{1}{2}t_\alpha^2)$$

when t is the *normal* critical level. The elimination of the Student distribution arises from the assumption of large n.

The derivation of this result hinges on the fact that $\sum_{i=1}^{n} x_i$ has for large n a normal distribution $N(\mu, \sqrt{n}\sigma)$. The central limit theorem ensures that this is approximately true whatever the distribution on an individual x_1 may be. Thus the stopping rule has a wider validity and may be used for all estimation processes so far discussed.

The expected sample size can be found to be

$$\frac{4\sigma^2 t_\alpha^2}{l^2}+\frac{1+t_\alpha^2}{2}$$

and the second term represents the additional sampling required to allow for the unknown value of σ.

If it is desired to quote a standard deviation for the estimate, the near normal distribution of the estimate can be used and the equivalent normal standard deviation is quoted

$$a=\frac{l}{2t_\alpha}$$

There is very little theory developed concerning the use of control variates to increase the precision of an estimate.

If the control variate is y with known mean μ and variance ω^2 and correlation ρ with z, the estimate is

$$\hat{z}_6=\bar{z}-\alpha\bar{y}+\alpha\mu$$

$$V(\hat{z}_6)=V(\bar{z}-\alpha\bar{y})=(\sigma^2-2\alpha\rho\sigma\omega+\alpha^2\omega^2)/n$$

This is minimised by choosing $\alpha=\dfrac{\rho\sigma}{\omega}$

Then $$V(\hat{z}_6)=\sigma^2(1-\rho^2)/n$$

The technique normally estimates α by least squares as

$$\hat{\alpha}=\frac{\sum_i y_i z_i-n\bar{y}\bar{z}}{\sum_i y_i^2-\bar{y}^2}$$

Now regard the set of sample values as a single multivariate sample. In repetition of sub-samples we are concerned with

$$E\{\hat{\alpha}(\bar{y}-\mu)\}=E\left\{\frac{\sum_i (y_i-\bar{y})(z_i-\bar{z})\sum_i (y_i-\mu)}{n\Sigma(y_i-\bar{y})^2}\right\}\neq 0 \quad \text{in general.}$$

Thus estimation of α from the data will introduce bias. This can be avoided by dividing the sample into two parts S_0 and S_1 containing n_0 and n_1 samples respectively.

Put $$\hat{\alpha}_k=\sum_{S_k}(y_i-\bar{y}_k)(z_i-\bar{z}_k)/\Sigma(z_i-\bar{z}_k)^2$$

where $$\bar{y}_k=\sum_{S_k} y_i/n_k \qquad \bar{z}_k=\sum_{S_k} z_i/n_k \qquad k=0,1$$

Now α_0 is independent of $\bar{z}_1, \bar{y}_1$ and α_1 is independent of $\bar{z}_0, \bar{y}_0$. Thus we may use the estimate

$$\hat{z}_7 = \tfrac{1}{2}[\{\bar{z}_1 - \alpha_2(\bar{y}_1 - \mu)\} + \{\bar{z}_2 - \alpha_1(\bar{y}_2 - \mu)\}]$$

Each of $\bar{z}_i - \alpha_{1-i}(\bar{y}_i - \mu)(i = 0, 1)$ is an unbiased estimate and so their mean is also unbiased.

Now
$$V(\hat{\alpha}) \simeq \frac{\sigma^2(1-\rho^2)}{n\omega^2}$$

from the theory of least squares.

After a little tedious algebra we obtain

$$V(\hat{z}_7) = \left[\tfrac{1}{4}2\sigma^2(1-\rho^2)\left(\frac{1}{n_1}+\frac{1}{n_2}\right) + E(\alpha_1\bar{y}_1)E(\alpha_2\bar{y}_2)\right]$$

$$= \tfrac{1}{2}\sigma^2(1-\rho^2)\left(\frac{1}{n_1}+\frac{1}{n_2}\right) + 0\left(\frac{1}{n_1 n_2}\right)$$

This is clearly minimised if $n_1 = n_2 = n/2$ and the variance becomes

$$2\sigma^2(1-\rho^2)/n$$

The price paid to remove the bias is quite high. A generalisation of this technique due to Tukey helps. Divide the total sample into k equal parts and estimate α from each of the samples formed by deleting one part at a time. If the parts are $S_1, S_2, \ldots, S_k$, define

$$\hat{\alpha}_j = \sum_{i \bar{\varepsilon} S_j} (y_i - \bar{y}_j)(z_i - z_j) / \sum_{i \bar{\varepsilon} S_j} (y_i - \bar{y}_j)^2$$

$$\bar{y}_j = \frac{k}{(k-1)n} \sum_{i \bar{\varepsilon} S_j} y_i \qquad \bar{z}_j = \frac{k}{(k-1)n} \sum_{i \bar{\varepsilon} S_j} z_i$$

$$\bar{\bar{y}}_j = \frac{k}{n} \sum_{i \varepsilon S_j} y_i \qquad \bar{\bar{z}}_j = \frac{k}{n} \sum_{i \varepsilon S_j} z_i$$

$$\hat{z}_8 = \frac{1}{k} \sum_i (\bar{\bar{z}}_i - \hat{\alpha}_i \bar{\bar{y}}_i)$$

Then to the same order of accuracy

$$V(\hat{z}_8) = \frac{k}{k-1} \frac{\sigma^2(1-\rho^2)}{n} + 0\left(\frac{k^2}{n^2}\right)$$

Clearly k cannot be made too large or the neglected term will become comparable with the main term.

The control variate technique involves more labour than simple sampling and there is a minimum correlation between y and z before the method is economic. Suppose the labour of the control variate technique is λ times that of the simple sampling procedure. In most cases $\lambda \geqq 2$.

The control variate technique is advantageous if

$$\frac{k}{k-1}\sigma^2(1-\rho^2)\frac{\lambda}{n} < \frac{\sigma^2}{n}$$

$$\text{i.e.} \quad \rho > \sqrt{\frac{k(\lambda-1)+1}{k\lambda}}$$

For the ideal case $\lambda = 2$, $k \to \infty$ the criterion becomes $\rho > \frac{1}{\sqrt{2}}$

Hammersley has suggested the use of negative correlation between two estimates as a method of reducing variation. Basically these techniques depend on the fact that if z_1 and z_2 are unbiased estimates of some parameter, then

$$z = a_1 z_1 + a_2 z_2 \qquad (\text{where } a_1 + a_2 = 1)$$

is also unbiased and the variance of z is given by

$$U(z) = a_1^2\sigma_1^2 + 2\rho a_1 a_2 \sigma_1\sigma_2 + a_2^2\sigma_2^2$$

This is minimised by choosing a_1 and a_2 as

$$a_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_2^2 - 2\rho\sigma_1\sigma_2 + \sigma_1^2} \qquad a_2 = \frac{\sigma_1^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_1^2}$$

With these values $U(z)$ reduces to

$$U_{opt}(z) = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2}$$

It can be shown that this is a monotonic increasing function of ρ and so the larger the negative correlation between z_1, z_2, the more accurate is the estimate z.

The usual and most important case is $\sigma_1 = \sigma_2$ when as would be expected

$$a_1 = a_2 = \tfrac{1}{2} \qquad U(z) = \frac{\sigma^2}{2}(1+\rho)$$

In this antithetic procedure, for a gain in efficiency we shall require

$$\lambda\sigma_1^2\sigma_2^2(1-\rho^2) < \min(\sigma_1^2, \sigma_2^2)(\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2)$$

Assuming $\sigma_1 \leqq \sigma_2$ this reduces to

$$\lambda\sigma_2^2\rho^2 - 2\rho\sigma_1\sigma_2 + \sigma_1^2 - (\lambda-1)\sigma_2^2 > 0$$

This is true for all $\rho < 0$ if $\sigma_1/\sigma_2 \geqq \sqrt{\lambda-1}$. (The simple case $\sigma_1 = \sigma_2$, $\lambda = 2$ satisfies this relationship.) If this is not so, then the equality can be reduced to a pair of linear inequalities. This gives the conditions

$$\rho > \frac{\sigma_1}{\lambda\sigma_2}\left[1 + \left\{(\lambda-1)\left(\lambda\frac{\sigma_2^2}{\sigma_1^2} - 1\right)\right\}^{\frac{1}{2}}\right]$$

$$< \frac{\sigma_1}{\lambda\sigma_2}\left[1 - \left\{(\lambda-1)\left(\lambda\frac{\sigma_2^2}{\sigma_1^2} - 1\right)\right\}^{\frac{1}{2}}\right]$$

The principal use of this technique has been with stratified sampling. If the strata are chosen so that $p_j = p$ const, then we may sample by taking a point in each stratum and repeating n/k times.

Consider a univariate case. Suppose as is usual that for each j, $p_j(x)$ is non-zero throughout the region R_j which now reduces to a simple interval. The estimate for $\bar{z}_j$ may be taken as $z(x)p_j(x)$ where x is uniform over the range of R_j.

Thus writing $f(x)$ for $z(x)p_j(x)$ and assuming that $[\alpha_j, \alpha_{j+1}]$ is the range R_j our estimate is

$$\sum_{j=1}^{k} f\{\alpha_j + \xi_j(\alpha_{j+1} - \alpha_j)\}$$

where ξ_j are uniform random variates. The normal stratified sampling procedure would take the ξ_j independent. The antithetic variate technique consists of taking the variate negatively correlated.

In the simplest case $k = 2$, the best possible relation is $\xi_2 = 1 - \xi_1$ which ensures a correlation of -1 in the ξ's but not of course in the two components of the estimate.

For larger values of k, the choice of a relationship is more difficult and in practice the technique resolves into using a power of 2 for k.

The whole process rapidly becomes similar in form to the ordinary methods of numerical analysis for evaluating the integral of a function as a weighted sum of ordinates.

The application of antithetic variates to other situations will be discussed in a later chapter.

CHAPTER 10

SIMPLE QUEUEING PROBLEMS

The sampling experiments discussed in earlier chapters have been applied to formal mathematical problems. We now turn to a study of more practical experiments concerned with industrial problems. The major difference here is that the statistics of interest are no longer derived as complicated functions of probability distributions but are expressions characterising a probabilistic process.

The simplest processes arising in industry concern queueing systems and in this chapter we illustrate various methods of conducting sampling experiments on such systems.

A simple queueing system consists of a collection of items to be processed, and a mechanism for processing them. To specify the system we must know:

(i) the time taken by the serving mechanism (the server) to process a given item;
(ii) the rules for the formation of the collection of items requiring service;
(iii) the rules for selecting an item for service by the server from the collection.

The time to service an item is usually variable and indeed it is this variability that raises the problem of the behaviour of the system. It is natural to describe this variability by a probability distribution of service times for the items. This still leaves open the relation between the service time of successive items. In the absence of definite evidence that one service can affect the next, independence is assumed and the set of service times is taken on samples of independent items from the *service time distribution.*

This assumes that the items serviced constitute a homogeneous set. It is possible to generalise the situation and stratify the items and associate a different service time distribution with each sub-set.

The collection of items requiring service varies with time and the rules for its change form the most convenient definition of the collection (the queue). After processing, items usually leave the queue. It remains to

specify the arrival of items to the queue. This is most conveniently done by specifying the times that items join the queue and, if there is stratification, what sub-set they belong to. These times do not usually form a regular sequence, i.e. the *inter-arrival intervals* vary. Once again, the most convenient description of these varying intervals is by a probability distribution. The question of the relation between successive intervals also arises and in the face of ignorance about the relation, successive intervals are assumed independent. However, if we are dealing with a succession of queues, the intervals between arrivals at one queue may be determined by the output of a preceding process.

Finally, the *queue discipline* must be specified. This consists of a rule for determining if the server is to serve an item and if so which item to serve. Usually, in the interests of high throughput, service is always given if there is an item in the queue to receive it. The choice of item usually depends on the order of arrival and the common rule is first come, first served. Special cases, such as railway wagons in a siding awaiting shunting, give rise to a last come, first served discipline. If the items are stratified, then the sub-sets may be priority-ordered with arrival time priority within sub-sets. In this case, the possibility of abandoning service of an item on arrival of a higher priority one arises, and then the rules for completion of the partially serviced item need specification.

All but the simplest queue systems defy analytic solution, although the mathematical theory of queues is a powerful tool for discovering the kind of phenomenon that can arise.

Once a complete description of the system is given it becomes possible to simulate it by sampling the appropriate distributions for process and inter-arrival times, and performing the necessary book-keeping. Conversely, the attempt to describe the procedure for simulating the system is often a valuable way of disclosing gaps in the complete description of the queue system.

We now turn to the problems of describing the results of simulating a queue system. The purpose of such a sampling study is to measure the congestion that the variable input and process time produce. The congestion can either be measured from the point of view of the system —how many items await service—or from that of the items—how long they wait. In the latter case, the natural description is to give the distribution of waiting times. The system description is more difficult as the number waiting at any instant depends on the number waiting at earlier times. Usually, the maximum queue that develops in a given period

is the number of principal interest, since space has to be provided for the items.

Suppose then that a simulation is required to study the maximum length of queue which develops in a given period, in a system consisting of a single service with a given service time distribution (R, say) independent of the item being served or of any history of service times, which is fed by a source of items arriving at intervals that have another given distribution (S, say). Suppose the system starts with an empty queue and an idle server.

All that is required is a count of the queue size as items enter and leave it. These changes are caused by events and a history of the events must be generated. As each item arrives, it is possible to predict the time of arrival of the next item by drawing a sample from S. Similarly, as each

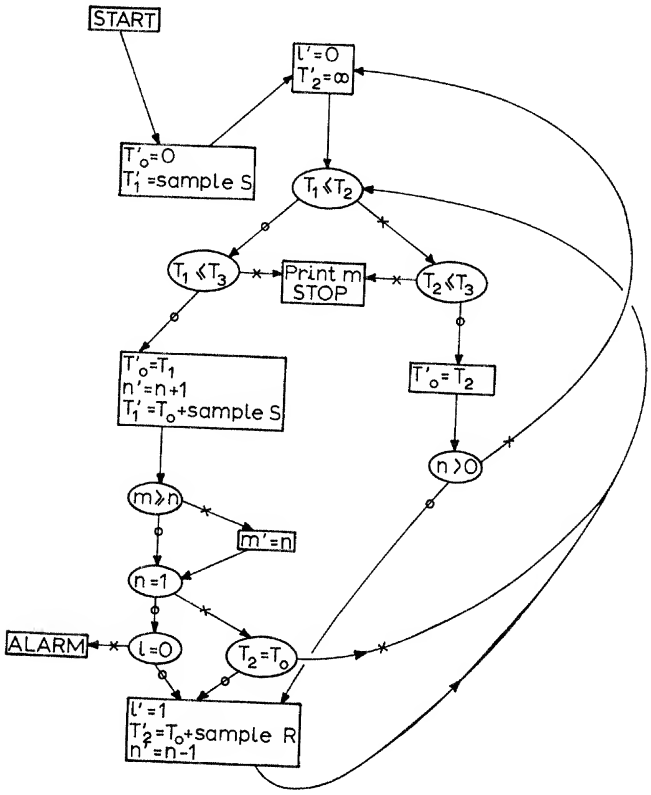


Figure 25

item starts processing, the time at which it will be completed can be predicted by drawing a sample from R.

At any stage, the next event to occur is found from the minimum of these two predictions.

Let the current time that has been reached in the simulation be T_0, the predicted next arrival T_1, the predicted end of service T_2 and the specified duration T_3. Let the current size of queue be n and use a marker l to denote if the server is active or not, set $l = 0$ if idle and $l = 1$ if active. Let the maximum queue size to date be m.

The flow diagram is then as shown in Figure 25.

In practice, the calculation would be repeated many times so that the statistical behaviour of the maximum queue size could be studied, and this flow diagram would be a sub-flow of a complete sampling procedure. For example, it may be proposed to accommodate a queue of size M. What is the frequency p of runs for which this is not large enough?

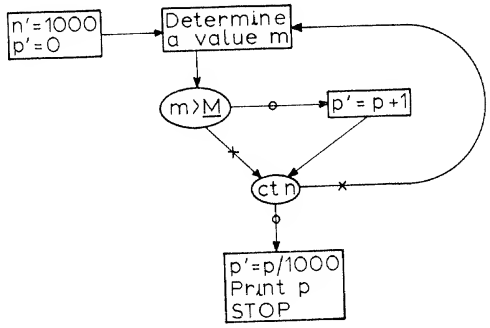


Figure 26

If in the practical problem M can be adjusted, it is more economic to form a frequency distribution for m and use this to estimate p for each of a range of values of M. This same function can be used to estimate the necessary M to ensure that the frequency of overload of the queueing area is limited to any particular figure.

The frequency being estimated is, by the nature of the problem, a low one and consequently will have a high variability, and thus will need a large sample to estimate it accurately. What can be done to reduce the sampling required?

First in the sub-flow diagram, simulation can be stopped as soon as m

reaches the value M (as the maximum in the range under consideration) since m is a monotonic function of the duration.

But a more powerful technique is to divide the total duration into two equal parts and perform a set of simulations over the first half, recording both the maximum and the final queues. These cases that have a large final queue are most likely to lead to large maximum queues in the second half of the duration and a second set of half duration simulations can be performed, using different starting conditions.

From the first simulations, estimates of frequencies $p(n)$ of different final queue lengths n are available and can be used to weight the conditional distribution of maximum queue length in the second duration.

In symbols, we take $p(m, n|r)$ to be the frequency of the system reaching a maximum queue length of m in a half duration, starting at length r and finishing at length n.

The distribution of maximum queue length m_1 in the first half duration is

$$p_1(m, -) \sum_n p(m, n|0)$$

and distribution of final length n is

$$p_1(-, n) = \sum_m p(m, n|0)$$

If the duration is long compared with the individual process times and inter-arrival intervals, the distribution of the maximum in the second half is nearly independent of that of m in the first half.

$$p_2(m, n) = \sum_r p_1(-, r)p(m, n|r)$$

and $$p_2(m, -) = \sum_r p_1(-, r) \sum_n p(m, n|r) = \sum_r p_1(-, r)p(m, -|r)$$

We may choose the amount of sampling for each of the starting conditions in the second half of the duration to increase the accuracy. An optimum rule is hard to formulate and sample sizes proportional to initial queue size are satisfactory.

Finally, the overall maximum is then found by sampling from the distribution $p_1(m, -)$ and $p_2(m, -)$ and forming the distribution of the maximum.

This technique is not quite accurate as the distribution $p(m, n, r)$ is not completely specified by the initial queue size r but also depends on the state of the server and the time of arrival of the first item.

In principle, the technique described for the single service queue of homogeneous items can be extended to several servers and a stratified population of items.

For example, suppose there is one server and two types of item; the server has a different distribution of process times for each type of item. Assume that the first type of item has complete priority over the second type. There are at least two possible ways of describing the input to the system.

The first assumes that the items of the two kinds arrive independently and each has its own characteristic inter-arrival distributions. The second assumes that the items arrive with a given inter-arrival distribution and on arrival are categorised as type 1 or 2 by random choice, the first item consisting of $p\%$ of the input and each choice being made independently of the previous choices. It can be shown that these two models only coincide if the inter-arrival distributions are exponential. Of course, both models are capable of several complications.

Suppose we make the first assumption. A possible flow diagram for a simulation to find the maximum queue developing is given in Figure 27.

Let T_0 represent the simulation time, T_1, T_2 the times of arrival of the next items of each kind; T_3 the time of end of service on the currently processing item by the server and T_4 the duration. Suppose n_1, n_2 give current size of the queues of the two kinds, n the total queue and m the maximum queue to date. Use a mark l to indicate the state of the server; $l = 0$, empty; $l = 1$, serving first type of item; $l = 2$, serving second type of item. Let the two process distributions be $R1$, $R2$ respectively and the two inter-arrival distributions be $S1$, $S2$ respectively.

The flow diagram based on this method will become even more complicated if the types of item are more numerous; if there is more than one server the network becomes almost impossible to draw.

However, there is a more systematic method of proceeding. Suppose there are k types of item with priority of service given by their type number. Let the times of arrival of items be T_i, $i = 1, \ldots, k$; T_0 being, as usual, simulation time. Let there be s servers each independently drawing from the queue. Suppose in the event of two servers competing for an item to serve that a similar priority system based on server number applies. Let the times of completion of service by the server be T_{k+i}, $i = 1, \ldots, s$. Let the service time distribution of the ith server serving the jth type of item be $R(i, j)$. Suppose $n_i(i = 1, \ldots, k)$ give the current sizes of queues of each kind of item, n the total queue

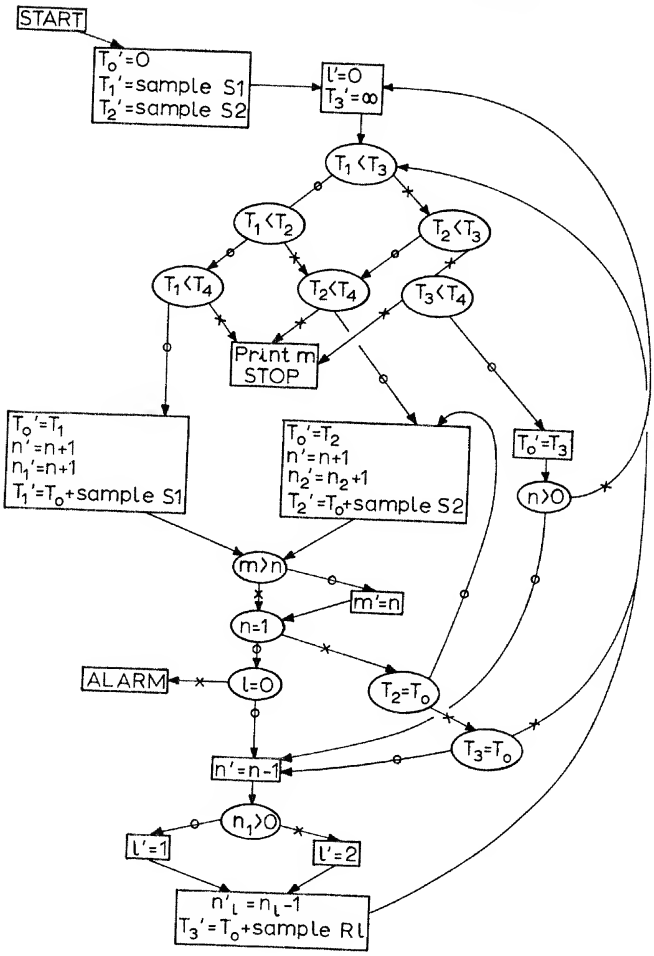


Figure 27

and m the maximum queue to date. Use marks l_j to denote the state of the servers. Let the inter-arrival distributions for the ith type of item be S_i, and the duration be T_{k+s+1}.

The flow diagram proceeds to determine the earliest time any event occurs and then searches in turn over the machines. For each possible event (service-end or item-arrival) which actually occurs, a sub-flow is entered. At the conclusion of each sub-flow the search is resumed.

The need to set the time of an idle server to a large value is

unnecessary, since in the time advancing search, only times greater than the present simulation time need be considered.

The flow diagram is as follows:

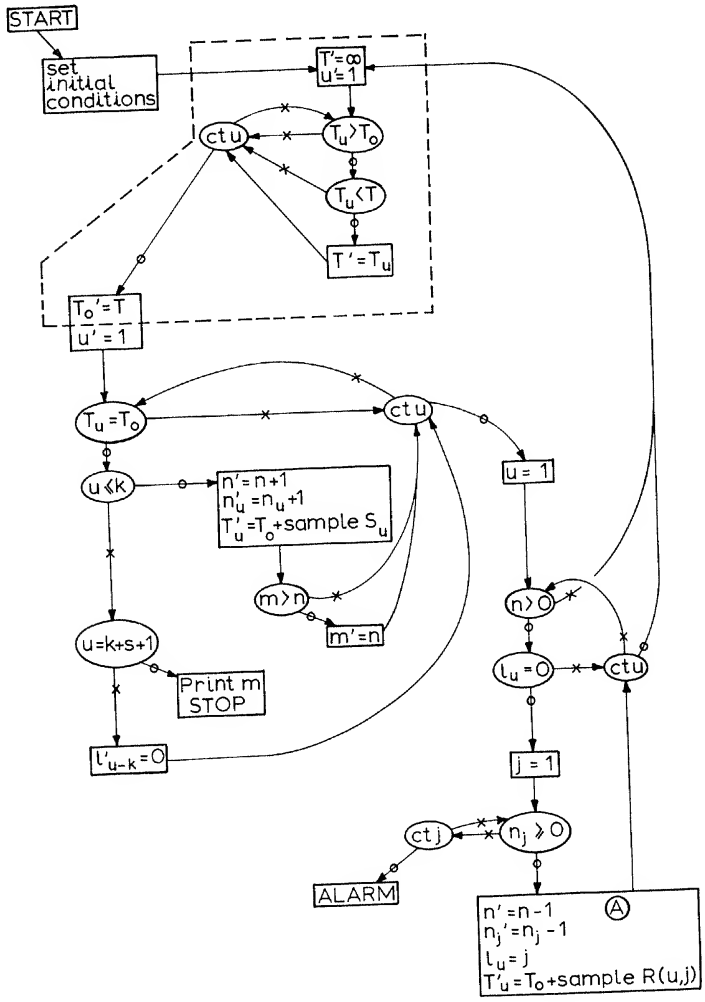


Figure 28

If the utilisation of the servers is of interest, this can be obtained from the waiting times and the flow diagram is easily adjusted for this.

Suppose the waiting times of the servers are being stored in W_u (initially zero) then the order

$$W'_u = W_u + T_0 - T_u$$

added to the sub-flow marked (A) will collect the waiting times and then the final printing must be adjusted to print these quantities. We shall find that the sub-flow enclosed in the dotted area is of constant occurrence and is abbreviated in later diagrams to ADVANCE T_0.

If only waiting times are required, we can proceed differently. First, consider a simple single service queue of one type of item. Suppose that the server has been waiting and has just received an item. Let the next item arrive at T_1 and let service be complete at T_2.

There will be waiting on the next item if $T_1 > T_2$ of duration $T_1 - T_2$. Otherwise there will be no waiting and the second next item will arrive at T'_1 say, and the processing of the next item will finish at T'_2. The second next item will cause a wait only if $T'_1 > T'_2$.

In this way we can determine the next waiting time, when it occurs, and the number of the items n processed since the last delay. It is more convenient now to limit the sampling to a fixed number of waiting times (held in a register m, say). Suppose the inter-arrival and process time distributions are $R1$ and $R2$ respectively. Then the flow diagram reads:

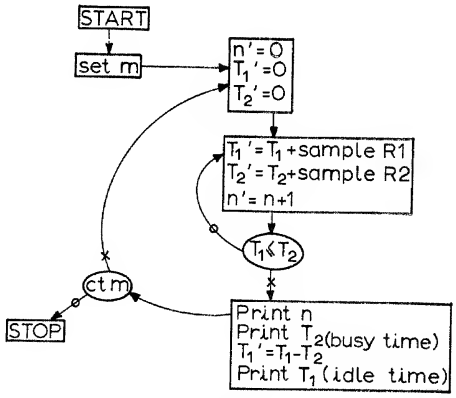


Figure 29

This technique is an example of a *recurrent event*—a concept introduced by Feller. We study the queue process only at times when a specified event occurs; in this case the event is the queue becoming empty. We are able, for this special state of the queue, to describe the

distribution of time between the occurrences of this state without studying all the intervening states. This arises because there is only one change of the system giving rise to the state of a zero queue. For all other queue lengths there are two ways to reach the length and the technique fails.

In particular, it is not possible to adapt this technique to find intervals between the time when the queue reaches a fixed maximum.

However, it is possible to generalise the waiting time formulation to deal with a stratified input. Suppose the items fall into k classes and, as before, initially that one item of class i, say, has just arrived and has been seized by the waiting server. Let inter-arrival distributions of items of class i be Si and assume that they all have common service time distribution R.

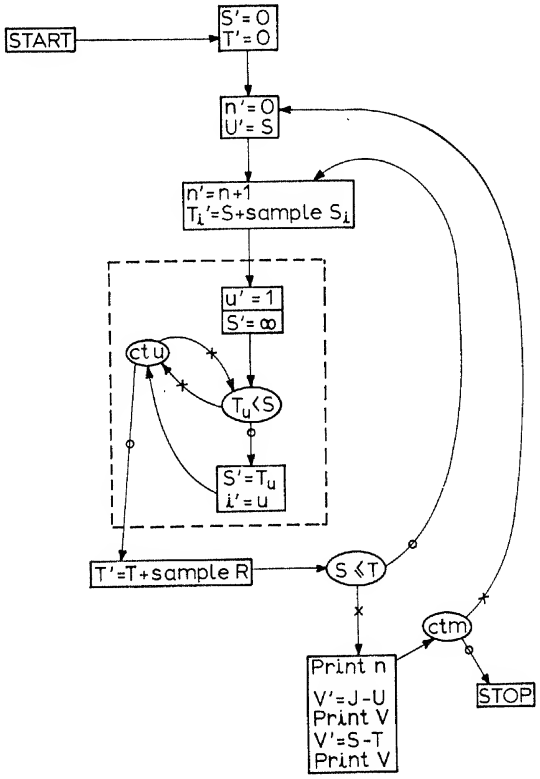


Figure 30

The sub-flow enclosed in dotted lines determines the minimum of a set and identifies an element attaining that lower bound. This process is so commonly required that the abbreviated expression

$$(i, S)' = \min_u (T_u)$$

will be used.

If several values attain the lower bound, this routine gives the one of lowest index number. By changing the inequality to $T_u \leqq S$, the index given will be the highest of those eligible. This will be denoted by

$$(S, i)' = \min_u (T_u)$$

Similarly, if a set of tests W must be satisfied before an item is admissible as a candidate in the search we write

$$(i, S)' = \min_u (T_u)/W$$

For example if a value W_u is associated with each machine and we require to determine the earliest-numbered machine with minimum value of T_u for those with non-zero W-values, we would write

$$(i, S)' = \min_u (S_u)/W_u \neq 0$$

The flow diagram for this is

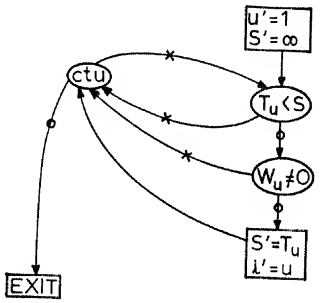


Figure 31

If there is more than one test in W these follow the / separated by commas, e.g.

$$(i, S)' = \min_u (T_u)/W_u \neq 0, S_u = 0, P_u > 0$$

A similar notation is used to determine the element maximising a quantity S

$$(i, S)' = \max_{u} (S_u)$$

In the present example, it is of no consequence which alternative minimisation is used. T is advanced for each of any set of items arriving simultaneously, before S is advanced.

If the server has a different process time distribution for each type of item, a slightly different procedure is necessary. We shall assume as before that the index priority system applies.

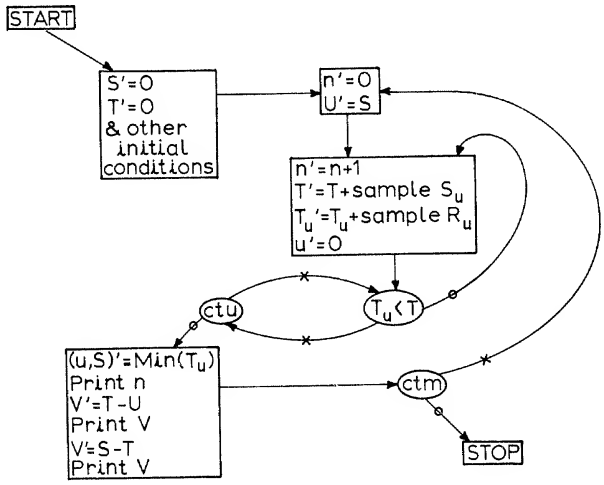


Figure 32

In this and the preceding example, the initial conditions are not completely specified, as at the end of the waiting period the items of the other classes must have their return times fixed. In practice, arbitrary but reasonable values can be assigned and the first few results discarded until steady conditions are established.

The problem of many servers to a single queue can be dealt with in a similar fashion. Now let T_u be times of return of the server ($u = 1, \ldots, S$), with process time distribution Ru, T the time of next arrival, inter-arrival distribution S. Figure 33 gives a possible flow diagram.

Finally, it is left as an exercise to the reader to develop the flow diagram for the completely general case of multiple servers of a stratified queue.

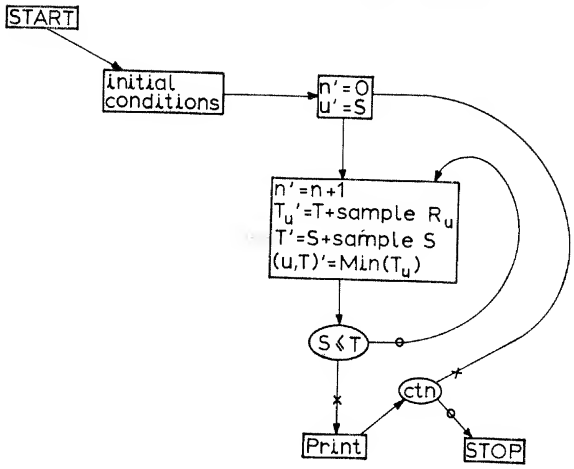


Figure 33

The analysis of queues has so far assumed an indefinite supply of items of process and this sampling has not depended on the number in the queue or the rate of processing.

Often the items processed involve a carrying mechanism that is held in the queue until the item starts processing or in some cases until after the process is finished. We shall consider the latter case. An example of this is a coal unloading machine, which empties wagons of coal into a stock pile and is then released to collect more coal. The number of wagons in the queue is limited to the total number of wagons and the arrival of wagons cannot then be described by a statistical distribution of inter-arrival times.

Suppose there are k wagons and each has a filling time after release by the emptying machine that can be described as a sample from a common filling distribution S. Suppose the emptying time is also described by a distribution R.

Let the number of wagons in the queue be Q, the time that the machine will finish unloading its present wagon (or finished unloading its last wagon if it is now idle) be T_{k+1}, the time that the ith wagon will arrive (or did arrive) at the queue is T_i, $(i = 1, \ldots, k)$ and finally the current time is T_0 and duration is T_{k+2}.

At any instant, the next change of the system will occur when a wagon arrives or the machine finishes emptying a wagon. If it is empty-

ing a wagon let S denote the number of the wagon. Let the queue discipline be first come, first served.

Suppose the required information is the maximum size of the queue during the run and a list of the individual waiting times of the wagons. Then the flow diagram for the simulation is given by

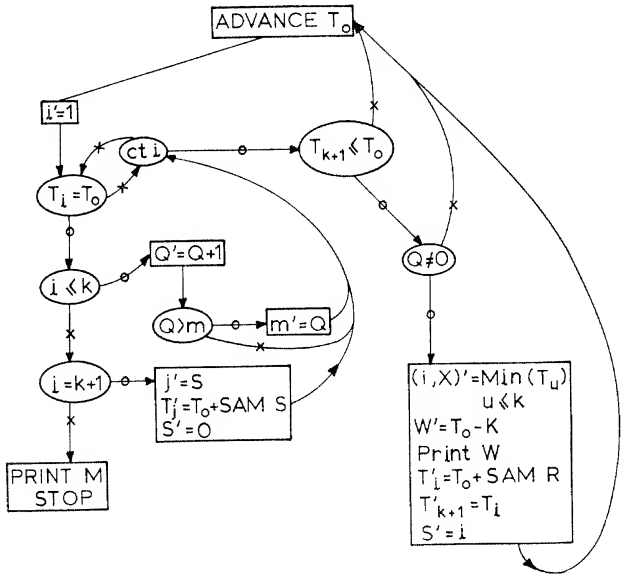


Figure 34

The method can clearly be extended to several machines serving a common queue of wagons. The machines should be ordered in priority to deal with the possibility of two machines becoming available for service simultaneously. Similarly if two wagons join the queue together, the one of lowest number will be chosen for service. More complicated selection rules can be arranged by extending the programme.

If only waiting times are required, then the recurrent event technique can be used. Suppose there are k wagons and m machines and the times of delivery of wagons are W_i, $(i = 1, 2, \ldots, k)$ and the times they finish a service are S_i. If service is in process on wagon i put $W_i = 0$ $(i = 1, \ldots, m)$. The objective is to measure the first m waiting times of machines and the interval between waits.

Suppose at time 0, a service has just started and all machines are

now serving, and there are no wagons in the queue, i.e. a wait by a machine has just finished.

Then the next machine to finish service will wait unless a wagon delivery has been made before the end of its service; if it does not wait then the next possibility of a wait is if the second wagon has not arrived before the second service finishes. This process can be repeated until a wait occurs.

Let the name of the wagon being serviced at zero time be held in V, then the programme reads

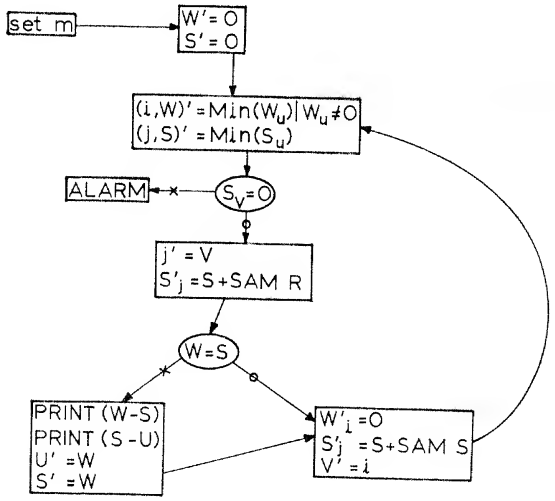


Figure 35

There is a similar solution to the problem of determining the waiting times of the wagons. This also starts from the condition where a service has just started and all items are busy and the queue is empty and we test that the next machine completes service before the next arrival. The details are left as an exercise to the reader.

Another variant of the queue problem concerns a multiple service queue in which arrivals select a queue by some rule and can then only be served by the machine associated with that queue. As a simple example, suppose the queue of minimum length on arrival is joined, that all machines have a common process distribution R and that the inter-arrival intervals are given by the distribution S.

If there are k machines whose end of service times are T_i and queues of length $Q_i (i = 1, \ldots, k)$ then with T_{k+1} for the next arrival time, T_0 and T_{k+2} as before, we have a programme to determine the maximum queue sizes $M_i (i = 1, 2, \ldots, k)$ as follows:

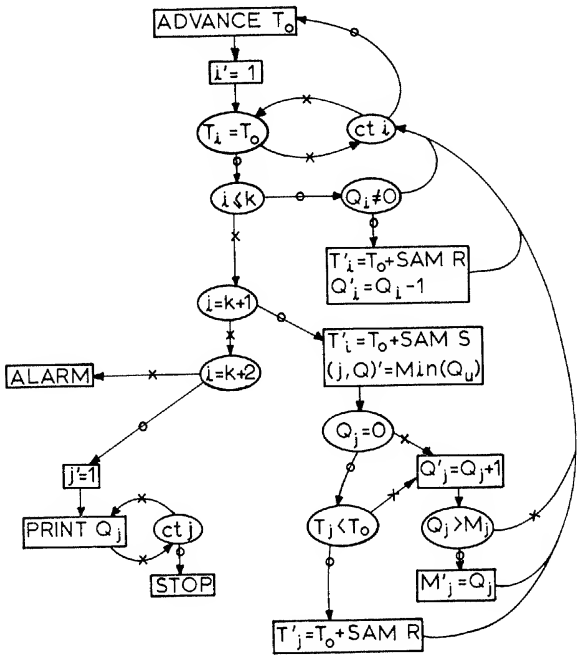


Figure 36

The ordering of machines is important in this example as the adjustment of queue lengths must take place before any item arriving at the same time makes a selection of the queue to join.

Finally, we consider the situation in which there is a set of queues fed serially from each other, the output of one forming the queue of the next. Input to the first machine is supposed to be described as usual by an inter-arrival distribution S and the machines have process time distributions $R_i (i = 1, 2, \ldots, k)$. It is convenient to change the notation and use T for current time, T_0 for the arrival events and T_{k+1} for duration. Set Q_0 to a negative value and put $R0$ for S. A possible flow diagram is given in Figure 37.

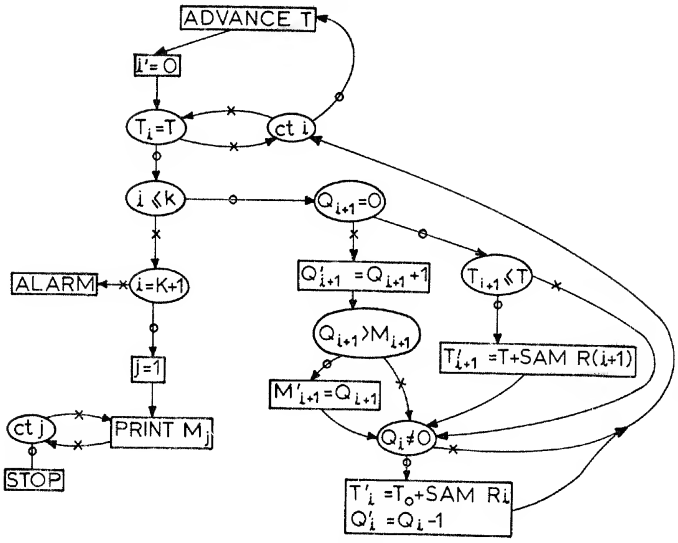


Figure 37

There are, no doubt, many other queueing systems that could be described, but the foregoing has indicated the type of attack that will enable a simulation of such systems to be written.

In the next chapter a different class of problem concerning queues is discussed and techniques for handling lists of customers are described.

CHAPTER 11

FURTHER QUEUE PROBLEMS

The analyses of the previous chapter have all been system or server oriented. We now turn to a customer oriented analysis and derive the history of each customer in the system. In general, this is much more complicated and involves keeping lists for each item in the queue. In the other analyses, the queue discipline, apart from the priority rule, has not been invoked. In a customer analysis, the discipline becomes all important.

However, the simple single server, unstratified, first come, first served queue system is amenable to a simple analysis. The relevant history of an item consists of the time of arrival and the times service starts and finishes. Let these be denoted by X_r, Y_r, Z_r respectively. Then we have the following recursive equations:

$$X_{r+1} = X_r + \text{sample } S$$
$$Y_{r+1} = \max(X_{r+1}, Z_r)$$
$$Z_{r+1} = Y_{r+1} + \text{sample } R$$

where R and S are the service time and inter-arrival distributions respectively.

In addition, the cumulative idle time W_r of the server can be obtained from

$$W_{r+1} = W_r + \max(0, X_{r+1} - Z_r)$$

The stratified queue can also be dealt with in this fashion if there is no priority system and the description of the arrivals follow the second form suggested in the opening section of the last chapter. The equations simply become in an obvious notation:

$$u = \text{sample of item type}$$
$$X_{r+1} = X_r + \text{sample } S_u$$
$$Y_{r+1} = \max(X_{r+1}, Z_r)$$
$$Z_{r+1} = Y_{r+1} + \text{sample } R_u$$
$$W_{r+1} = W_r + \max(0, X_{r+1} - Z_r)$$

For any other queue discipline, we must revert to the event type of approach.

Consider the single server queue with a stratified queue and a queue discipline based on priority and first come, first served, within a stratum. As in the system-oriented case described earlier the times of arrival of items of the different kinds and the time of service are required. In addition, each item in the queue needs historical information about it storing in a list.

The list will be arranged to accommodate a maximal number of items. This can be fixed either by guess-work or employing a system oriented simulation to determine the maximal number that will be in the queue and then using the same pseudo-random number starting values to ensure generating the same sample times. In the former case steps must be taken to stop the simulation if the guess proves wrong. The positions in the list are numbered 1, 2, ..., m and each is used to store

U_s type of item
X_s time of arrival
Y_s time of service start
Z_s time of service end.

If a position is not in use $U_s = 0$

The marker l associated with the server is now used to hold the list number s of the item under service. If $l = 0$ the server is idle. Using the remaining notation as before, we can draw the flow diagram to print the history of each item when service is complete, as in Figure 38.

The sub-flow enclosed in dotted lines occurs frequently and can be abbreviated to

LIST ENTRY $(U, S)' = (u, T_0)$

This technique has the disadvantage that the whole list is searched each time an item is required, and the items could have been stored in separate sub-lists on entry. This raises the problem of the necessary length for each sub-list and the tendency to use more storage to ensure that no sub-list overflows. This extra storage required is usually an overriding objection to the multiple list arrangement.

The alternative left is to sort the list into a priority order. For this to be effective, extensive book-keeping of the break points in the lists becomes necessary. We can arrange for each section to contain the items in any order or we can maintain a strict ordering by time of arrival. We consider the latter scheme first.

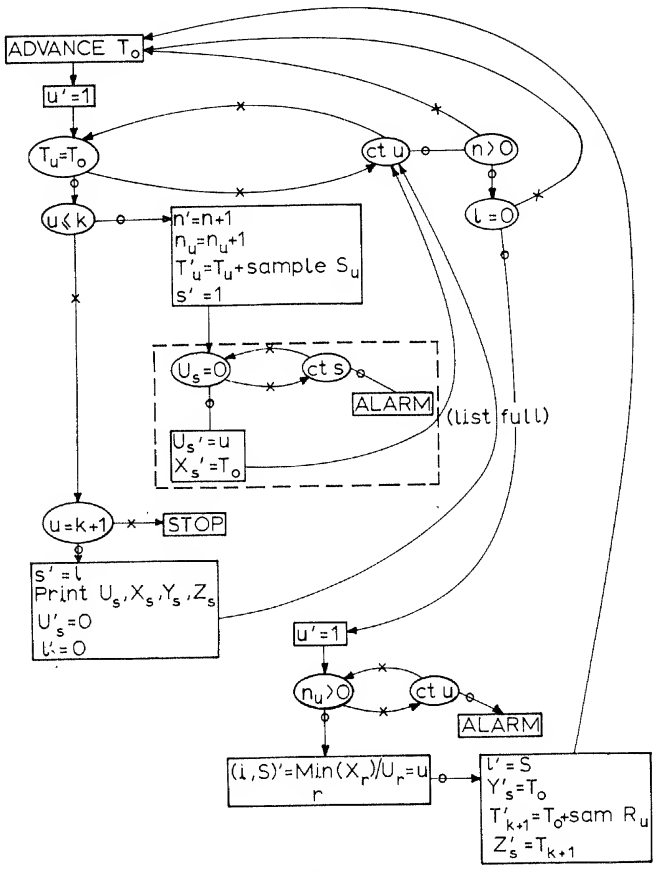


Figure 38

Let the register M_u contain the name of the leading item in the uth sub-list. We use a set of registers U, X, Y, Z to hold the information obtained from a selected item in the list. The process

$$U' = U_s, \qquad X' = X_s, \qquad Y' = Y_s, \qquad Z' = Z_s$$

is abbreviated to 'READ s'. The reverse process is abbreviated to 'ENTER s'.

Withdrawing a uth type item from such an ordered list is merely

$$s' = M_u$$

$$M'_u = M_u + 1$$

READ s

$U_s' = 0$

For entering an item, a second temporary register is desirable. Reading and writing to this are abbreviated to READ s and ENTER s. It is also necessary to interchange the contents of the registers, which is abbreviated to SWITCH.

The process of reading an item from one class creates a space in the sub-list of the next higher class. The first step in entering an item is to see if such a space (or spaces) exist(s). Otherwise the items in the list must be moved down until the last displaced item can be entered in an empty position. This may not happen before the bottom of the list is reached, but there may still be room in an earlier part of the list. Thus it is convenient to regard the list as circular, the last position being followed by the first. In this displacing process the registers M_u must be adjusted.

A possible flow diagram is given below:

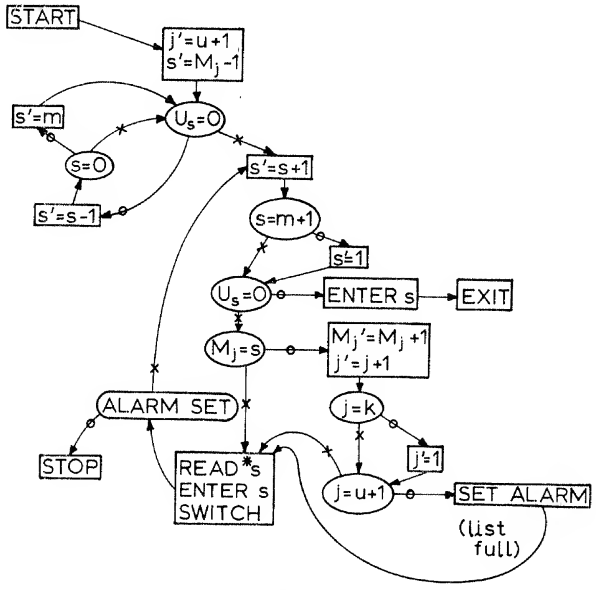


Figure 39

An alternative method of proceeding in this case of ordered sub-lists is to divide the available storage into units and use ordered aggregates of these to form a sub-list. As the items in any sub-list increase, unused

units are added to the aggregate and as items are used, emptying a unit, this is released for use by other sub-lists.

This method is of value only if the total number of items likely to be stored is very large, since, if there is to be little wastage, each unit must be small compared to the total storage, and as each unit requires space to describe how it is being used, each unit must be large compared to the number of variables used to describe its use.

There are two variants of the method. The first regards the units as items in a set of ordered sub-lists. To enter an item, the appropriate sub-list of units is read to determine the last item in the sub-list. This gives the name of the unit to enter the item. This unit is scanned and, if any space is left, the item inserted in the next space in the unit. Otherwise the sub-list of empty units is consulted to find an empty unit and the item inserted into the leading space; at the same time the unit is deleted from the sub-list of empty items and entered on the sub-list of units allotted to this appropriate category.

To remove an item from the list, the appropriate sub-list of units is consulted and the leading item noted. This gives the name of the unit whose leading item is the required item. The item is removed from the unit and if this is now empty, the leading item in the sub-list of units is removed and entered as the last number in the sub-list of empty units.

A flexible storage arrangement could be used for the sub-lists, but the complexity of the process in this case would be unlikely to be worth the storage saved, bearing in mind that the orders to execute the arithmetic of the process also take storage space. Using fixed length sub-lists for the names of units, the storage arrangement is illustrated diagrammatically in Figure 40.

The unit lists are on the left giving for each of three types of item the unit number and an associated serial number to give order of entry. The units holding the items are displayed on the right and the chains linking them into a complete list are shown.

To enter an item in type 1 sub-list, the units sub-list is scanned for maximum serial number; in this case 8. Unit 20 is given as the unit containing the tail of the list. This unit is now scanned to find the first empty cell; in this case cell 4.

If an item is to be entered in the type 2 sub-list the unit 18 is selected since its serial number is 4, the maximum in that sub-list. On scanning this sub-list it is found to be full. The first non-empty cell of the empty units list gives unit 17. The item is entered in the first cell of this unit and the first empty cell in the type 2 sub-list is found. This is cell 6,

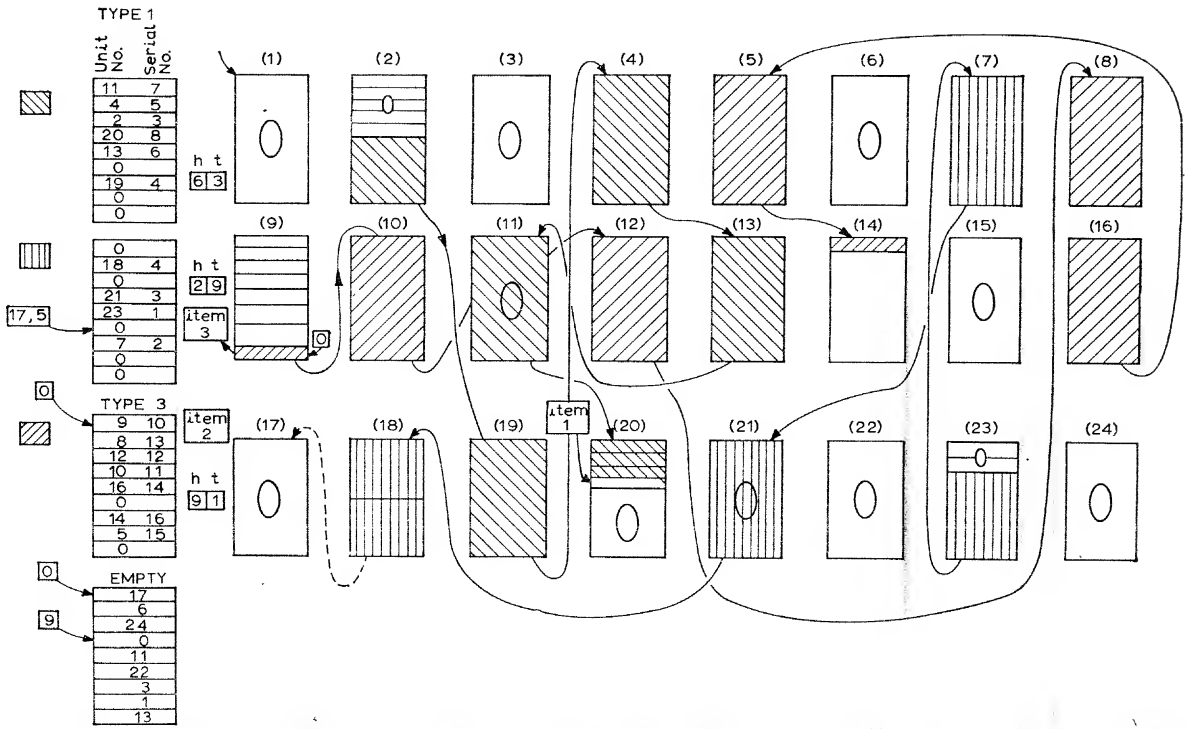

TYPE 1
Unit No.
Serial No.
11 7
4 5
2 3
20 8
13 6
0
19 4
0
0
h t
6 3
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
0
18 4
0
21 3
23 1
0
7 2
0
0
h t
2 9
item 3
17,5
(9)
(10)
(11)
(12)
(13)
(14)
(15)
(16)
TYPE 3
9 10
8 13
12 12
10 11
16 14
0
14 16
5 15
0
item 2
h t
9 1
(17)
(18)
(19)
item 1
(20)
(21)
(22)
(23)
(24)
EMPTY
17
6
24
0
11
22
3
1
13

Figure 40

in which we enter the pair (17, 5). The cell containing 17 in the empty sub-list is cleared.

To remove an item, the unit sub-list is searched for minimum serial number. For a type 3 item this gives 10, corresponding to unit 9. Unit 9 is searched for the first non-zero item, and this is found to be last. Removing this item leaves unit 9 empty. Thus the cell serial number 10 is cleared and an empty cell in the empty list is found and 9 inserted.

In practice, the scan of the units can be avoided by keeping a record of the position of the head and tail of each list. Let there be k items in each unit list with entries (U_i, S_i) $i = 1, \ldots, k$; h, t, the ordinal number of the head and tail of the list. Each variable requires a further suffix j to denote the sub-list number, but this will be omitted without loss of generality. The empty list is unordered and is denoted by (V_i).

The item in position m of unit l is denoted by $I_{l,m}$. Let X, x, i, j be working spaces. The flow diagram for an entry of an item I is then given by:

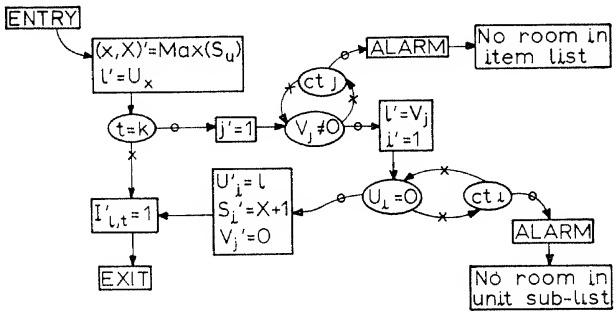


Figure 41

Similarly for a withdrawal we have:

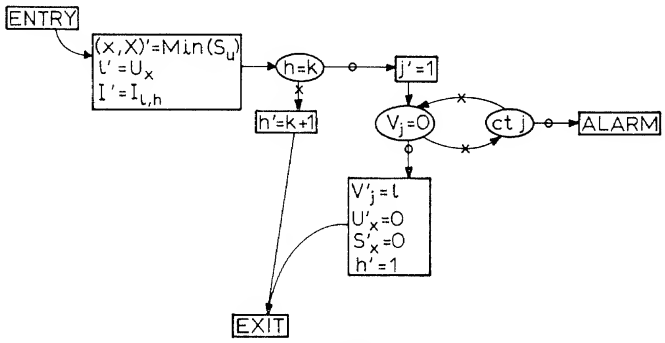


Figure 42

The second variant uses the principle of link chain storage. Instead of storing unit numbers and serial numbers in a two variable item list, a cell is associated with each unit and the name of the next higher and next lower unit stored in its two halves. Associated with each type of item is a 4-part cell containing the name of the unit and position in that unit of both the head and the tail of the list.

Suppose the two-part cell associated with a unit l is (U_l, V_l) and the 4-part cell associated with a given type is (i, h, j, t) while the first and last units in the chain of empty units are (x, y) $(U_i = V_j = U_x = V_u = 0)$.

The entering and removing of an item then becomes:

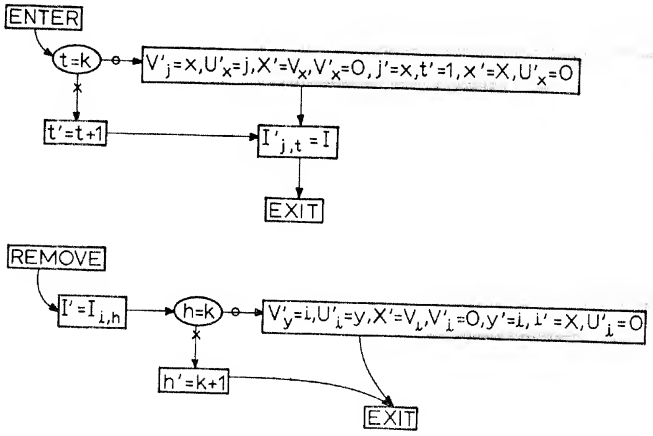


Figure 43

For the example of the figure the chain cells L_j are as follows:

1	(3, 13)		9	(0, 10)	[13, 0]	17	(0, 6)	[18, 0]
2	(0, 19)		10	(9, 12)	[0, 12]	18	(21, 0)	[21,17]
3	(22, 1)		11	(24, 22)		19	(2, 4)	
4	(19, 13)		12	(10, 8)		20	(11, 0)	
5	(16, 14)		13	(1, 0)	[1, 9]	21	(7, 18)	
6	(17, 24)	[0, 24]	14	(5, 0)		22	(11, 3)	
7	(23, 21)		15	(6, 17)		23	(0, 7)	
8	(12, 16)		16	(8, 5)		24	(6, 11)	

An arbitrary chain of empty units has been assumed. Thus for entry of a type 2 item we have

$$(i, h, j, t) = (23, 2, 18, 9) \qquad (x, y) = (17, 13)$$

We have $L'_{18} = (21, 17)$ $L'_{17} = (18, 0)$ $L'_6 = (0, 24)$

$(i, h, j, t)' = (23, 2, 17, 1)$ $(x, y)' = (26, 13)$

For the removal of a type 3 item $(i, h, j, t) = (9, 9, 14, 1)$

$L'_{13} = (1, 9)$ $L'_9(13, 0)$ $L'_{10} = (0, 12)$

$(i, h, j, t)' = (10, 1, 14, 1)$ $(x, y)' = (6, 9)$

This chain method is superior to the list of units method in the amount of storage involved and the amount of calculation involved, and will well repay study to understand the mechanism.

It may be used to deal with items if the amount of storage for each item required is high. Each item now has a link store and the 4-part cell reduces to a two-part one as for the empty units chain. In fact, the units of the former analysis are regarded as items. The simplification of the calculation is left to the reader.

We now turn to the problem of maintaining unordered sub-lists in a common list. This arises if the criterion for selecting an item from a sub-list is different from the first come, first served basis. For example, the item selected may be that with the earliest delivery promise.

Reverting to the notation of the first sub-list technique (p. 138) the selection of an item of type u is achieved by a scan of the sub-list where the start and finish cell names are M_u and $M_{u+1}-1$ respectively.

To enter an item I of type u we first see if it can be inserted at the end of the present sub-list. If not, it is entered in place of the head of the next sub-list $(u+1)$ and the process repeated to insert the displaced item into sub-list $(u+1)$. This continues until the item can be inserted or the whole list has been scanned. In this latter case an alarm is given.

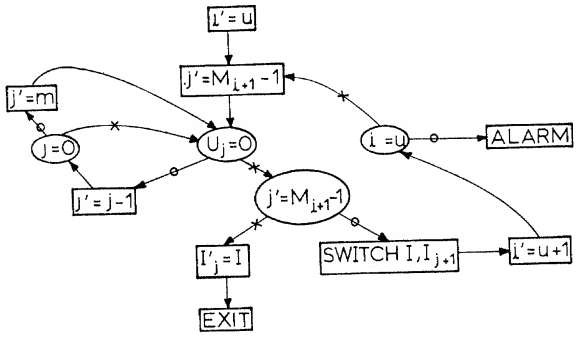


Figure 44

It is possible to allow free use of the list for storing items of all types and at each scan to record the first and last items of the type being selected and to use these values to determine the range of next scan. The attempt to enter a new item should start at the first item of the same kind and if an empty cell is not found before the last item of that kind, the last item record is adjusted. This assumes a circular list.

However, this method involves the storage of two references for each sub-list and will always involve larger scans than the first method.

If the storage capacity allows a second record for each sub-list, a better procedure is to record the location of the last item in each sub-list in N_u, say.

Then the insertion of an item of type u reduces to:

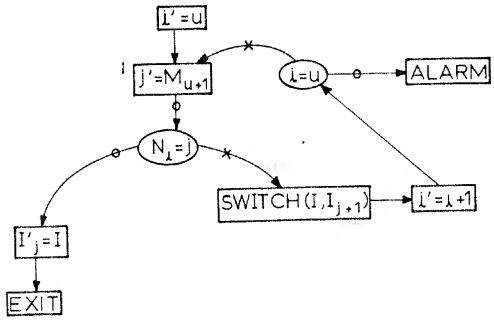


Figure 45

These techniques for handling lists can also be used to deal with items passing through a queueing system. The sub-lists then correspond to items in the different queues. By rearranging the items, searches for them when service can be given can be made over a reduced range.

Lists also arise in the study of other industrial simulations. This forms the subject of the next chapter.

CHAPTER 12

GENERAL SIMULATION PROBLEMS

Since the popularisation of queue theory, it has become common practice to describe an industrial plant as an aggregate of inter-connected queues. The theory that gave birth to this view of a plant is not, however, capable of offering a solution to many of the problems of industrial congestion and simulation is frequently used to study such problems. Earlier chapters have shown various techniques for simulating queue systems of varying complexity, and the extension to more complex systems can be envisaged.

However, the queue description does not meet all industrial situations. The most important exception is when machines undergo cycles of activities and co-operate in some (but not all) of their phases with other cyclic machines. These, in turn, co-operate with each other in other phases and so on. The time taken in the various phases are random variables and so machines do not always reach the co-operating phases together. This induces a delay in one of the machines. The whole situation can be likened to a set of stochastic gear wheels clicking round irregularly.

As a simple but illuminating example, consider an electric steel-making furnace. Such a furnace goes through a cycle of five phases: first, it is charged with pig iron and steel scrap; second, this is melted; then it is refined until the content of trace elements such as carbon have reached the required level; the resulting steel is poured (or tapped) into a ladle, and finally the furnace lining is repaired by charging a refractory material which replaces that lost by chemical erosion during the cycle. This last phase is known as 'fettling'.

The ladle used for tapping is held by a crane which removes it when full to a 'teeming' platform where it is poured into moulds. These moulds are mounted on a railway bogie which, after standing, is hauled by a locomotive to a stripping bay where the moulds are removed from the now solid steel. After stripping, the locomotive returns the empty moulds for refilling. Similarly, the ladle crane takes the empty ladle for cleaning before it is used for tapping again.

In practice, the basket, crane and loco cycles are fast enough to ensure

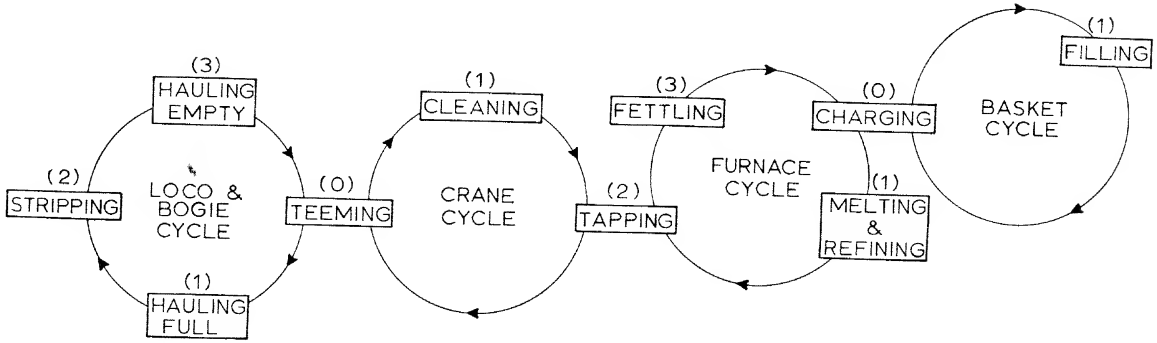

(3)
HAULING EMPTY
(2)
STRIPPING
LOCO & BOGIE CYCLE
(0)
TEEMING
(1)
HAULING FULL
(1)
CLEANING
CRANE CYCLE
(2)
TAPPING
(3)
FETTLING
FURNACE CYCLE
(0)
CHARGING
(1)
MELTING & REFINING
(1)
FILLING
BASKET CYCLE

Figure 46

that very little interference takes place, but often these machines serve several furnaces, when the congestion problem can arise. It is instructive to imagine that the various cycle times are similar so that the technique for simulating such a system can be developed. The complication of several furnaces and possibly several cranes and locos can be introduced later.

The process can be described in the diagrammatic form of Figure 46.

Each machine has a set of states corresponding to its various phases and distributions to describe the time spent in these phases. We must now introduce some notation.

Let the machines be numbered 1 for the basket, 2 for the furnace, 3 for the ladle and 4 for the loco. Let T_i be the time the ith machine will finish its current phase (or if it is idle the time it finished its last phase); S_i be the state of the machine (or if it is idle its last state). Number the states as follows:

Basket		*Furnace*		*Ladle*		*Loco*	
Charging	0	Charging	0	Teeming	0	Teeming	0
Filling	1	Melting and Refining	1	Cleaning	1	Hauling full	1
		Tapping	2	Tapping	2	Stripping	2
		Fettling	3			Hauling empty	3

The activities of melting and refining can be coalesced as the furnace moves from the first to the second state without invoking any other machine. This also avoids the difficulty that the refining time may be partly dependent on the melting time.

Let the time distribution of the ith machine in the jth state be $R(i, j)$; T_0 be the current time, T_5 be the duration. Let S_5 be a dummy variable always taking the value 0. Suppose we wish to measure the times between successive casts of steel being stripped.

Then the programme must arrange to change the state and time associated with each machine as it reaches the end of each phase. The activities of charging, tapping and teeming will demand special treatment as in each case two machines must have completed their preceding phases.

The first attempt at the flow diagram might appear as shown in Figure 47.

This diagram is seen to be very repetitive but not quite systematic; some of the exits to the '*ct i*' routine do calculations, others do not. However, it is possible to determine for each machine as its state is changed whether or not a state change is automatic on leaving that state.

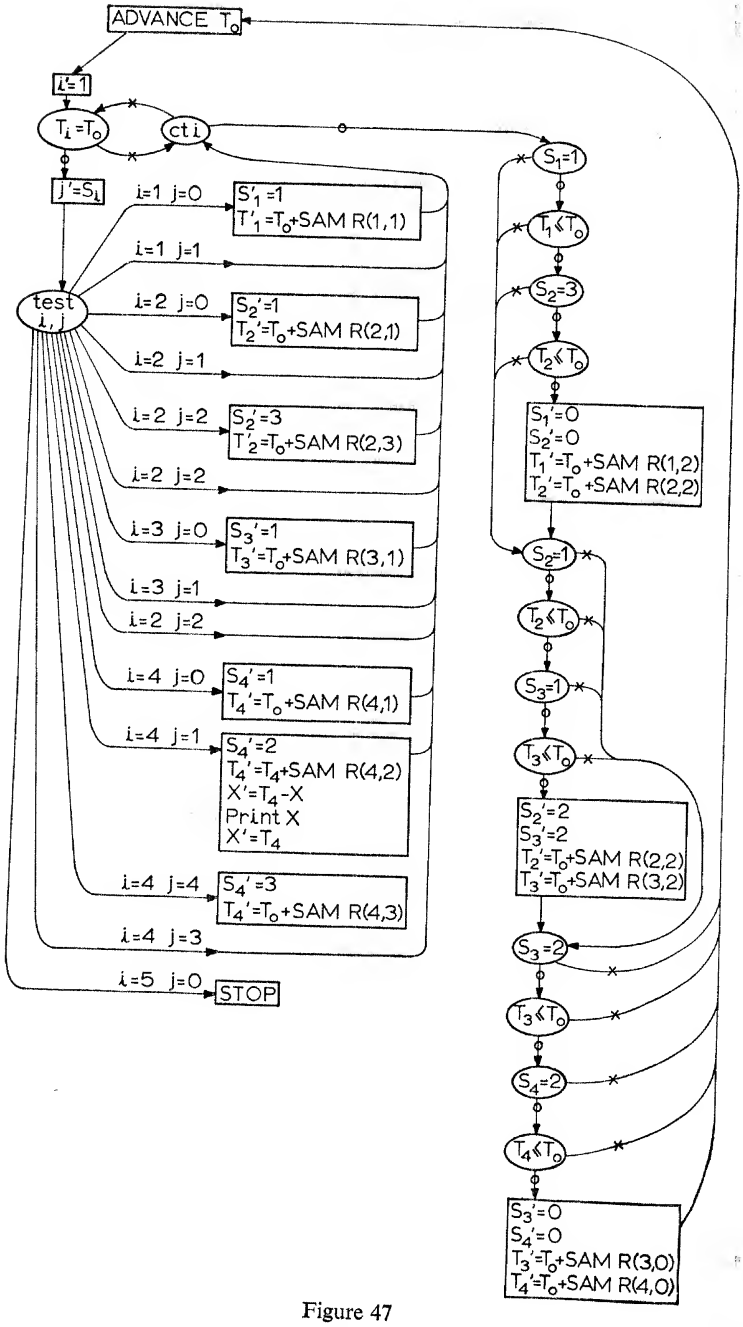

ADVANCE T_o
$i'=1$
$T_i=T_o$
ct i
$j'=S_i$
test i, j
i=1 j=0
$S'_1=1$
$T'_1=T_o+SAM\ R(1,1)$
i=1 j=1
i=2 j=0
$S'_2=1$
$T'_2=T_o+SAM\ R(2,1)$
i=2 j=1
i=2 j=2
$S'_2=3$
$T'_2=T_o+SAM\ R(2,3)$
i=2 j=2
i=3 j=0
$S'_3=1$
$T'_3=T_o+SAM\ R(3,1)$
i=3 j=1
i=2 j=2
i=4 j=0
$S'_4=1$
$T'_4=T_o+SAM\ R(4,1)$
i=4 j=1
$S'_4=2$
$T'_4=T_4+SAM\ R(4,2)$
$X'=T_4-X$
Print X
$X'=T_4$
i=4 j=4
$S'_4=3$
$T'_4=T_o+SAM\ R(4,3)$
i=4 j=3
i=5 j=0
STOP
$S_1=1$
$T_1<T_o$
$S_2=3$
$T_2<T_o$
$S'_1=0$
$S'_2=0$
$T'_1=T_o+SAM\ R(1,2)$
$T'_2=T_o+SAM\ R(2,2)$
$S_2=1$
$T_2<T_o$
$S_3=1$
$T_3<T_o$
$S'_2=2$
$S'_3=2$
$T'_2=T_o+SAM\ R(2,2)$
$T'_3=T_o+SAM\ R(3,2)$
$S_3=2$
$T_3<T_o$
$S_4=2$
$T_4<T_o$
$S'_3=0$
$S'_4=0$
$T'_3=T_o+SAM\ R(3,0)$
$T'_4=T_o+SAM\ R(4,0)$

Figure 47

Thus, for example, as the basket enters the state of charging, it can be predicted that, at the end of charging, filling will take place.

The required systematisation can be achieved by introducing some more constants. Let C_i be the cycle length of machine i and define b_{ij} as one if machine i changes to its next state on leaving the state j; and zero otherwise. For machine 5, put $b_{5,0} = 2$. Introduce variables B_i to hold the current values of b_{ij} for each machine.

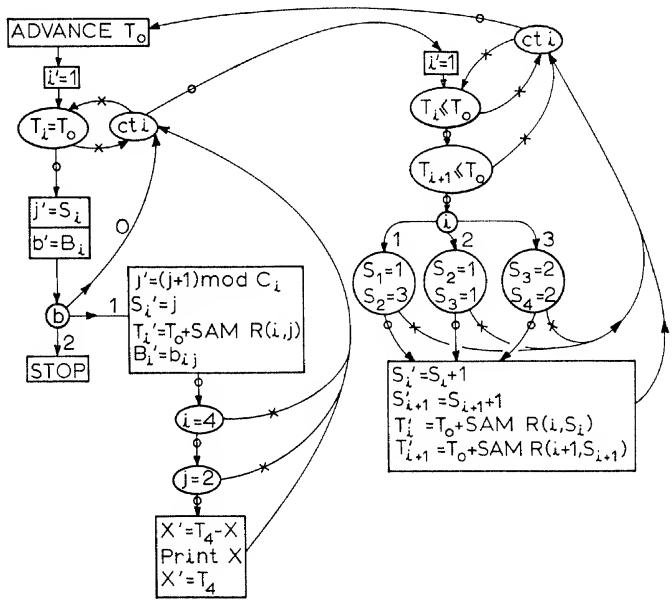


Figure 48

As a second problem, consider the problem of simulating the passage of traffic across time-sequenced traffic lights at a junction.

Since the example is for illustrative purposes, we will assume that there is no entry to the side road so that we do not need to consider the blocking effect of traffic turning right into the side road. The lights go through a cycle first allowing the main traffic to flow (state 0), halting all traffic (state 1), allowing the side road traffic to enter the main road (state 3) and finally halting all traffic again before re-starting the cycle. Assume that the times in these states are fixed at t_0, t_1, t_2, t_3 respectively.

The rules for describing the passage of traffic across the junction

must allow for the fact that vehicles which cross from rest will take longer than those which arrive when there is no queue at the lights.

The proposed rule divides the passage into two parts: in the first the vehicle moves from the queue into the junction and then in the second actually crosses. As soon as one vehicle has moved into the junction, another can start to move. In this way a continuous stream of traffic across the junction is achieved. If a vehicle arrives when there is no queue and it can cross, the first phase takes place instantaneously. Otherwise, the time a vehicle spends in the first phase is described by a distribution R, common for each of the three streams of traffic. The time in the second phase cannot be so described since the vehicles must leave the junction in order of arrival at the queue. To achieve this, a fixed time t_4 will be used. Vehicles arrive at the traffic streams with inter-arrival times given by distributions S_1, S_2, S_3.

The rule for the passage of traffic is simply that it can cross if and only if the lights allow and there is no vehicle on the junction.

Let the current time be T_0, the current (or immediate past) first phase crossings for the three streams end at times T_1, T_2, T_3 and the next arrivals of vehicles be at T_4, T_5, T_6. Finally, let T_7 be time of the next change of the light, T_8 the time the crossing will be clear and T_9 the duration. Let S denote the state of the lights and U_1, U_2, U_3 be the variables associated with the three streams of traffic taking the value one if the traffic stream can pass and zero otherwise. Let the queues in the three streams be Q_1, Q_2, Q_3. Stream 3 is chosen as the side road traffic and the purpose of the simulation is to measure the maximum queue developing in each stream.

A possible scheme is shown in the following flow diagram (Fig. 49).

There are some points of special interest in this example. The three times T_1, T_2, T_3 are reset each time a new vehicle is taken from the queue by the routine labelled A. When the queue empties, these times will lag behind T_0 and the routine A will not be entered again. Vehicles can never *start* moving from a queue by A; this is achieved by routines B and C which switch the U variables to allow traffic to move at the later of (i) a favourable light change and (ii) the junction becoming clear of traffic and, if there are vehicles queueing, resets T_1, T_2 or T_3 and causes the junction to become busy with the first vehicle crossing.

Vehicles which arrive when there is no queue proceed straight to the junction and only the junction clear time needs adjustment.

The adjustment of the U variable is most conveniently done as shown, since if it is transferred to the right-hand side, two versions are necessary,

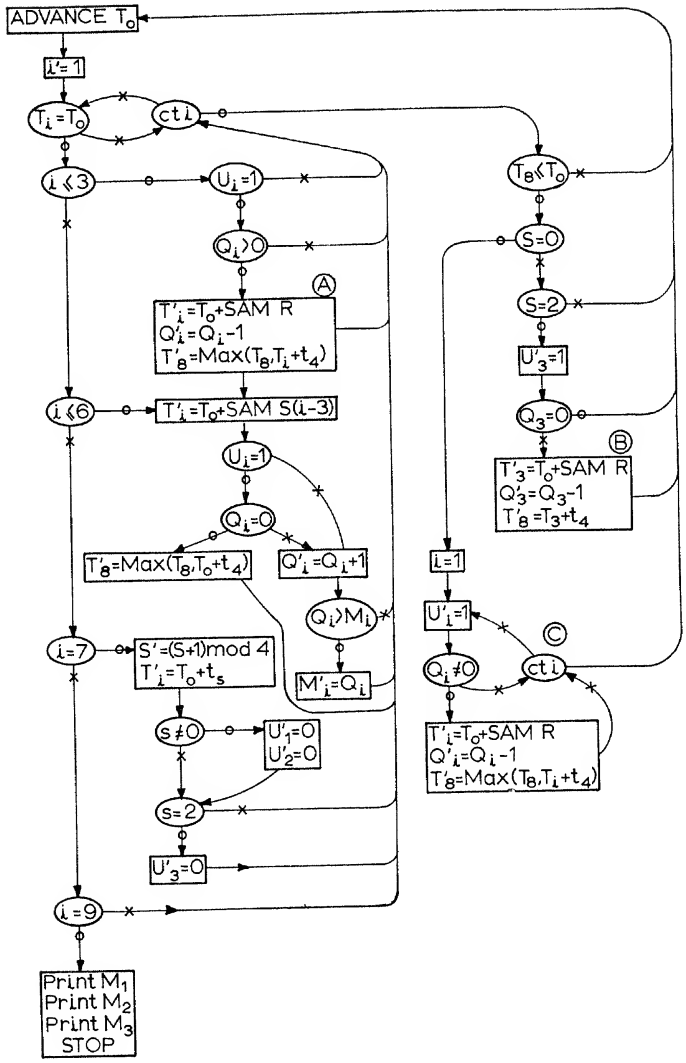


Figure 49

one to adjust them when the lights become favourable if $T_8 \leqq T_0$ and the other when the junction clears ($T_8 = T_0$), if the lights are favourable.

In this case the present numbering of the times is essential since the

adjustments to T_8 must be made prior to changing the U's. This ensures that 'tail-end Charlies' do not get hit by a new traffic flow.

Our third example concerns a machine shop machining foundry pieces. We assume, say, 3 machines served by one overhead crane which fetches the pieces from the foundry, helps with handling each piece part way through each processing and finally takes the machined pieces to a store. Suppose the loading time distribution is R_1, the first process time is given by distribution R_2, the handling time (shared with a crane) has distribution R_3 while the second process time and the unloading time are given by distributions R_4 and R_5 respectively. Finally, R_6 is the fetch-and-carry time distribution.

A priority rule must be given for the crane in the event of competition. This is simply that handling will take precedence over unloading. The crane always reloads a machine as its next job after unloading it. We assume the purpose of the study is to estimate the idle time of the crane. The whole process is illustrated below for one machine.

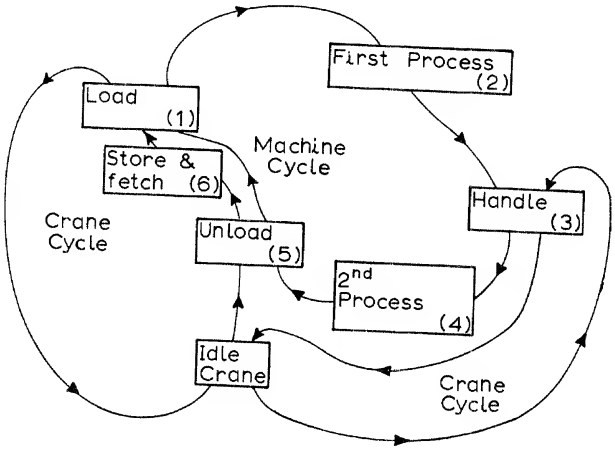


Figure 50

We assign variables as follows:

T_0 current time.

T_1, T_2 and T_3 the times the machines finish (or finished) their current phase.

T_4 the time the crane finishes its current job.

T_5 duration.

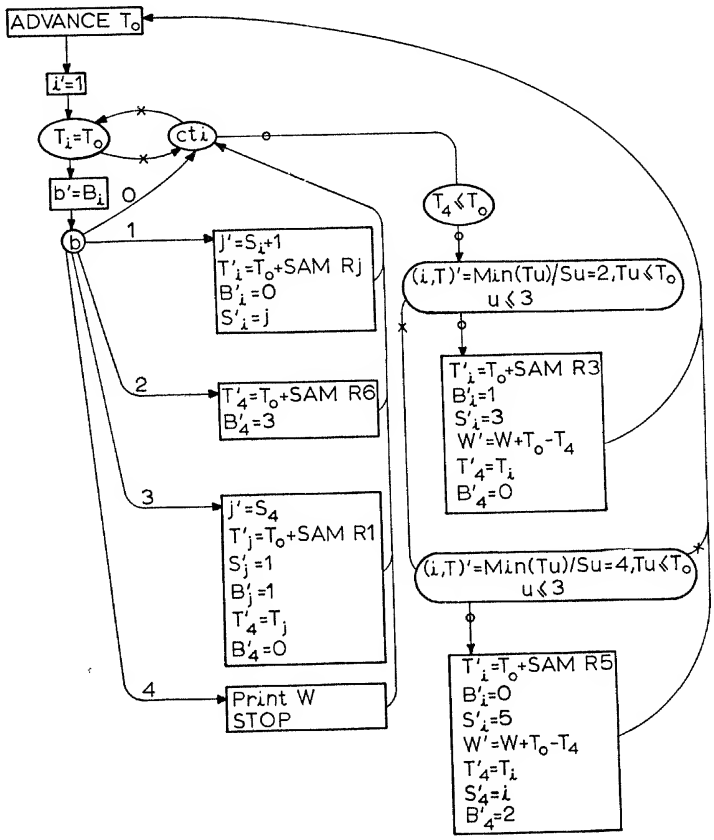
S_1, S_2, S_3 the states of the machines.

$S = 1$ Loading
$S = 2$ First Processing
$S = 3$ Handling
$S = 4$ Second Processing
$S = 5$ Unloading.

S_4 the name of machine which the crane is currently serving.
W the crane waiting time.
B_1, B_2, B_3, B_4, B_5 organisational storage

Put $B_5 = 4$.

The flow diagram for a simulation is given below:



Here the use of the organisational variables has been taken a stage further and used systematically to ensure that the correct adjustments are made as each event occurs.

The searches for the machine to service when it is known that a crane is available deserves mention. If no machines satisfy the side conditions, the cells i, T will not be set and the alternative exit is taken to by-pass the adjustments corresponding to the handling or unloading activities. If the first search is successful it is clear that the second cannot be so since T_4 is now greater than T_0.

Looking at these three very different examples as well as the later examples of queue systems, it can be seen that a general pattern emerges from them all. This can be expressed in the following flow diagram:

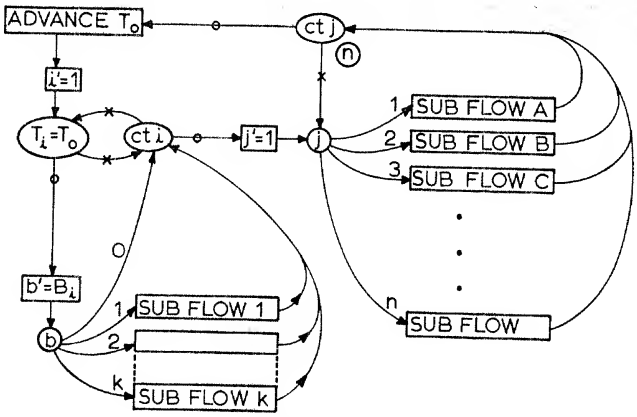


Figure 52

The difference of roles played by the numbered and lettered sub-flows deserves comment. A numbered sub-flow is executed at a fixed time, i.e. immediately T_0 reaches a time associated with an event. If the sub-flow is prefaced by tests which fail, the executive part of the sub-flow will not be entered at any other time until some other sub-flow changes a time or b-number. The purpose of the lettered sub-flows is to just so reinitiate the numbered sub-flows and to start activities which involve the occurrence of otherwise independent events. Examples of the use of lettered activities are:

(i) processing an item on a machine involves both an idle machine and a non-empty queue,

(ii) handling a piece involves both the machine having reached the stage where handling is required and that a crane is available.

From this general flow diagram and the examples which led to it, certain useful concepts can be defined which help in the description of flow charts. First, with every time variable, we associate a machine. Often this machine has a real existence, but sometimes an imaginary machine must be invented. Thus in the road traffic example there are imaginary vehicle loaders which take vehicles from a queue into the junction and an imaginary machine to measure the occupancy of the junction.

The times associated with these machines define events of some kind, and often we associate this with a change of state of the machine. We are thus led to the concept of machine-states, which can be an arbitrary assignment of numbers to the different conditions of the machine. The phases of the furnace of the first example constitute a typical example. There is no restriction to a single variable for the specification of the state of a machine.

When a machine i is made to start some activity, its state takes on a new number j, say, and the time T_i is given by adding the process time to T_0. The pair of instructions

$$T_i' = T_0 + t$$
$$S_i' = j$$

are spoken of as *committing* machine i to state j for time t; this latter is usually provided by a sampling process, but this is not necessarily so, e.g. the traffic light machine of the junction example. The test that a machine is *committed* is that $T_i > T_0$. If this is not so (i.e. $T_i \leqq T_0$) the last activity has been completed and the machine is idle, or to use a less flavoured word, is *available*.

A machine which has only one job, e.g. a server in a queue, need not have a state variable associated with it since the working state is equivalent to being committed and the idle state is equivalent to being available.

The various sub-flows adjust the states and times of machines to correspond to various activities on the plant and can conveniently be described as *activities*. The two types must be distinguished and the notation *B*-activities and *C*-activities is introduced. *C*-activities are conditional ones describing co-operative activities. *B*-activities are performed *immediately* a machine becomes available and often contain book-keeping instructions (hence *B* for book-keeping). The instruction to obey such a *B*-activity on becoming available is associated with a machine and the machine is said to be *engaged* to the activity. Several

machines may at any time be engaged to the same activity, in which case the activity usually involves the machine name as a variable index. A variable index b_i is associated with each machine to hold the name of the B-activity required. By convention zero will be used for the null activity.

With this structure understood the specification of a simulation becomes the specification of the machines, their states and initial conditions and the individual activities. Thus the simple single server queue simulation becomes in an obvious notation:

Machine	*Name*	*States*
1	Arrival machine	—
2	Server	Q the size of the queue M maximum size of queue

Initial conditions $T_1 = 1$, $b_1 = 1$, $Q = M = 0$, $T_2 = 0$, $b_2 = 0$.

*B*1. *Item arrives*

Increase queue by one.

If queue now exceeds maximum adjust the latter.

Commit the arrival machine for another fetch-and-carry time.

Engage to *B*1.

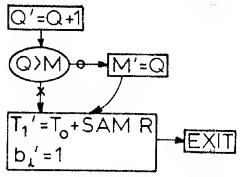


Figure 53

In practice, since the engagement *B*1 is not disturbed this does not require renewal in *B*1 and may be omitted.

*C*1. *Item served*

If there is a real queue and the server is available.

Commit the server for a service and reduce the queue by one.

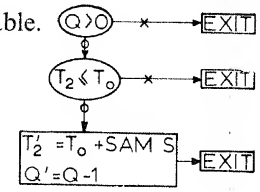


Figure 54

It is tempting to represent every real machine in a plant by a time-state pair of variables, but this is not always necessary or desirable. Each machine so introduced lengthens the process for determining the time of the next event and if a group of service machines are involved, all offering similar service, a C-activity will make a search over these machines to find one available.

To illustrate the point, suppose we are concerned with a set of nine lorries operating a cycle of arriving at a warehouse, waiting to be loaded by one of the five fork-lift trucks, travelling to a site and returning empty. The obvious procedure to determine the lost time of lorries waiting to be loaded is as follows:

Machine	*Name*	*States*
1–9	Lorries	S, 0 loading, 1 travelling
10–14	Trucks	—

Initial conditions — W(total waiting time) = 0.

*B*1. *Lorry Journeys*

Commit lorry to a journey.
[Mark lorry as travelling.]
Disengage lorry from *B*1.

*C*1. *Loading*

Set to first crane and lorry.
Find a lorry available [which has finished travelling].
Find a crane which is available.

Record waiting time of lorry.
Commit crane and lorry for loading time.

[Mark lorry as loading.]
Engage lorry to *B*1.
Return for other lorries.

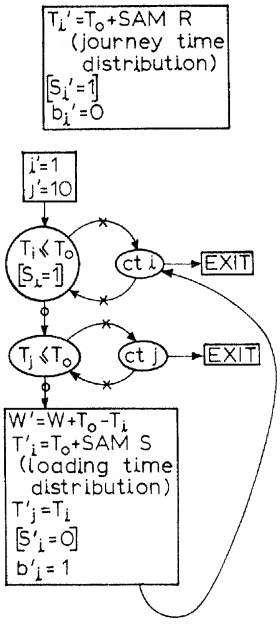


Figure 55

In *C*1, the return to the '*ct i*' instruction after committing the loaded lorry and crane is to ensure that all idle cranes and returned lorries are mated at the proper time.

It is not necessary in fact to use a state for the lorries since the only use for this is in the first test in *C*1. If a lorry is available at the time this activity is executed, it must be waiting for a load since loaded lorries are dispatched by *B*1 automatically before *C*1 is reached. Thus the expressions in square brackets can be omitted without loss.

However, a greater economy can be achieved by noting that since

the trucks are used only for one purpose, it suffices to keep a count of the trucks available. This gives a simulation as follows:

Machine	*Name*		*Initial value*
1–9	Lorries	Waiting time W	0
		Available trucks Q	5

*B*1

Increase count of trucks available. Commit lorry for a journey time. Disengage from *B*1.

$Q' = Q + 1$
$T_i' = T_o + SAM\ R$
$b_i' = 0$

*C*1

Set to first lorry.

If there is at least one truck available, find a lorry which is available.

Record the waiting time.
Reduce count of trucks available.
Commit lorry for a loading time.
Engage lorry to *B*1.

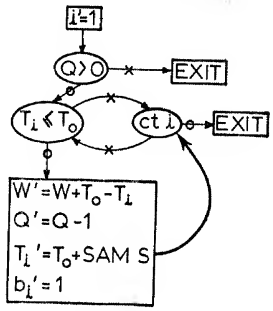


Figure 56

If a wheel diagram is drawn it is seen that the truck cycle has only one element in it. This is an indication that a mere count of the trucks available will suffice.

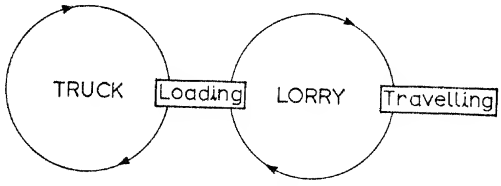


Figure 57

Similarly the type of activity required (*B* or *C*) is revealed by the diagram. Loading must be a *C*-activity since it involves the co-operation of machines (even if these are represented by counts) but travelling can be a *B*-activity as it occurs unconditionally after loading.

We now consider two further examples. The first is a simple queueing process. Items join a queue with a given distribution of inter-arrival

times, to await processing by a machine. The items are dealt with singly and the process times are independent samples from a given distribution. After processing, the items enter a stock. The stock is dispatched at regular intervals in batches of m items. If the stock is less than m items, no dispatch is made.

Write a simulation to determine the maximum queue which develops, the idle time on the machine and the number of missed dispatches, given 100 items in stock initially, an empty queue, an idle machine, a dispatch interval of 50 units, a duration of 500 units, and a batch size of 35.

Make an allocation as follows:

Machine	*Name*	*States*	*Initial Conditions*
1	Arrival machine	—	$T_1 = 1$ m/c arrives immediately
2	Process machine	Q queue	$T_2 = 0$ m/c idle
			$Q = 0$
		U cumulative idle time	$U = 0$
		M maximum queue	$M = 0$
3	Dispatch machine	S items in stock	$T_3 = 1$ dispatch starts immediately
			$S = 100$
		W missed dispatch count	$W = 0$

Also set initially $b_1 = 1$, $b_2 = 0$, $b_3 = 3$, $b_4 = 4$.

Use distributions $R1$ and $R2$ for inter-arrival times and process times respectively.

*B*1. *Arrivals*

When the arrival machine returns, sample a fetch-and-carry time, commit the machine to fetch-and-carry, increase the queue of items at the process machine by one and if the queue is now larger than the current maximum, adjust the latter. The initial engagement to $B1$ still holds and this machine repeats this cycle indefinitely.

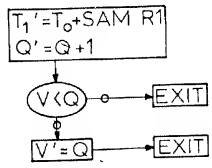


Figure 58

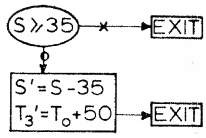
*B*2. *Increase Stock*

When an item finishes being processed on the process-machine, increase the stock by one item. The machine remains available until *C*2 recommits and engages it.

S'=S+1

*B*3. *Dispatch*

When the dispatcher returns for the next delivery, test if there are at least 35 items in stock and, if so, reduce stock by 35, and commit the dispatcher until the next dispatch time. Otherwise do nothing. *C*1 will then count the delay and dispatch the dispatcher.



*B*4. *Printing*

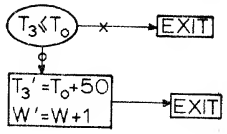
When the printing machine arrives at the end of the run, print the results.

PRINT M
PRINT U
PRINT W

Figure 59

*C*1. *Delays in Dispatch*

If the dispatch machine is available (which can only happen if *B*3 failed to dispatch 35 items), count a delay and commit the dispatcher until the next dispatch time.



*C*2. *Reduce Queue*

If the process machine is available and there is a queue of items for it, sample a process time for next item, add idle time to the cumulative total, remove one item from the queue and commit the machine to process the item.

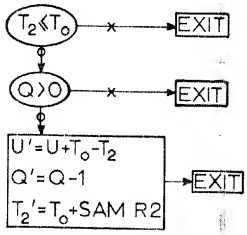


Figure 60

The ordering of the machines is relevant to the success of this model. *B*2 is entered before *B*3, since the machines involved are in that order. The stock is correctly increased before the test for 35 items is made.

The second example concerns a circular bus route which is served by 8 buses each able to carry up to 56 passengers. There are 12 bus stops around the route, and the length of each passenger journey is an independent sample from one distribution of journey lengths. Passengers join queues awaiting buses at each stop with inter-arrival times which are independent samples from distributions particular to each stop. The length of time any bus takes to travel from one stop to another is an independent sample from a distribution of stop-to-stop times.

The disembarkation and embarkation times for passengers at each stop are calculated as 5 units for the first passenger and an extra unit for each successive passenger.

Write a simulation to determine the maximum queue which is likely to develop at each bus stop.

Machine	*Name*	
1–8	Buses	U_i the next stop for the bus i P_i the number of passengers on the bus i Q_i the number of passengers joining the bus i at the current stop
9–21	Stops (Passenger arrival machine)	Q_i actual queue at the stop (i–8) M_i maximum queue at the stop (i–8) $P_{(r,s)}$ number of people on the rth bus due to alight at the sth stop W working space

Let the inter-arrival distribution for ith stop be R_i and the travel time distribution to stop j be S_j. Let the journey distance distribution be J.

Engage all stops to $B1$ ($b_9 = b_{10} = \ldots = b_{21} = 1$) and all travelling buses to $B2$. Enter $B3$ on duration.

B1. Bus Stop Queue

Add passenger to queue. Reset loading time for first passenger. Commit arrival machine to next arrival.

Adjust maximum queue if necessary.

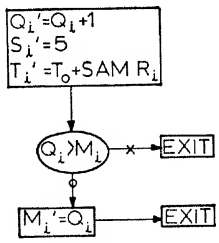


Figure 61

B2. Disembark

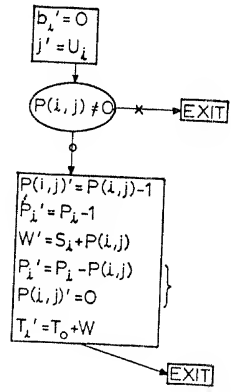
Disengage machine.
Pick up stop number.

If there are passengers on this bus to alight at this stop,

remove first passenger,

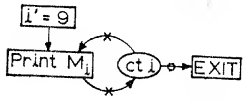
calculate alighting time,
adjust passenger lists and commit the bus to unload.

Figure 62



B3. Printing

Print the values of $M_9 \ldots M_{21}$.



C1. Load Passengers

Find an available bus, with empty seats, which is at a bus stop with a non-empty queue.

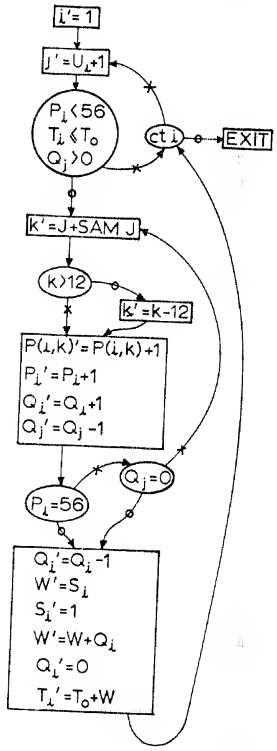
Sample a journey distance (number of stops travelled) and determine the destination of the passenger.

Adjust the count of passengers on the bus alighting at that stop; the number of passengers on the bus and the number embarking at the stop.

Repeat until the bus is either full or no queue is left.

Remove first passenger.
Allow time S to load passengers.
Adjust S for 'late-comers'.
Calculate rest of loading time.
Reset count of passengers embarking.
Commit bus.
Return for other buses, if any.

Figure 63



C2. Starting

Find a bus which is available (but with no passengers to load).

Calculate the next stop and commit bus for journey.

Engage to *B*2.

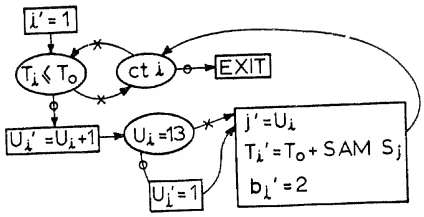


Figure 64

In this example, the ordering to the *C*-activities is important since buses which are available after *C*1 must be ready to depart. If the order were reversed, extra tests would be needed to establish that it was ready to depart.

The entities called machines need not be conventional machines. A powerful variant of the technique is available if a known group of items are to be processed by a series of machines. Each item undergoes a series of process $1, 2, \ldots, j, \ldots, k$ using one of n_j machines of type j for the jth process which takes a time given by a distribution R_j.

The m items to be processed are called machines $1, \ldots, i, \ldots m$ and each has a state associated with it—the stage of processing reached. Suppose we require the intervals between items leaving the system, assuming that items enter the system in order and are dealt with at each process in order.

Machines	*Name*	*State*
1–m	Items	$S\ j$: during jth process or waiting for $(j+1)$th process
		Q_j number of jth process machines available
		W last time item left system
		M count of items

All machines are engaged to *B*1.

*B*1

Set j to state of returning item.

Find an item which is available in state $j-1$, if any.

Commit it to process j. If no such item, increase count of available machine of type j.

If last process, print interval and test for last item.

Otherwise, test if any machines available for next process. If so, reduce count of such machines and commit item to the process.

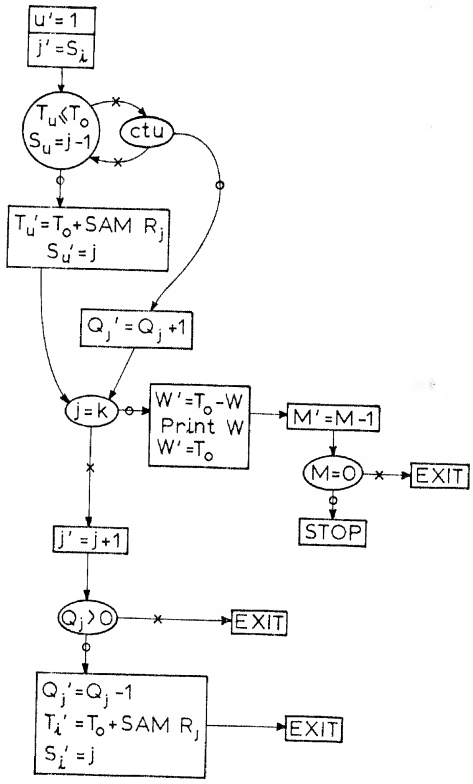


Figure 65

No *C*-activities are required since the *B*-activity looks back to see if the present process has a queue (formed by earlier processes being completed when no machines were available) as well as forward to see if the next process can start.

By keeping further counts C_j, say, of the queue sizes, the unsuccessful searches can be avoided.

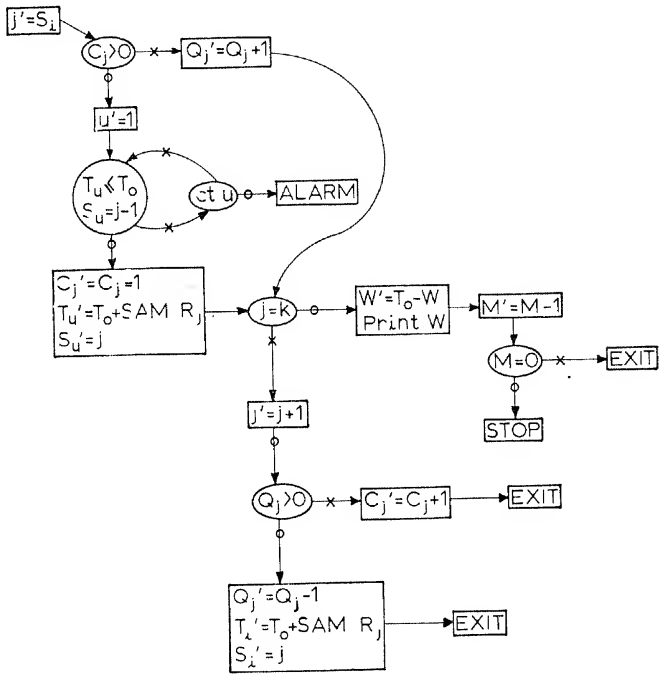


Figure 66

If some of the processes involve, in addition to the basic processing machine, an ancillary machine, such as a crane, and a group of these are available for common service, then this simple model breaks down. However, it is still possible to retain the same basic structure.

Associate with each process a constant k_i taking value one if a crane is needed, zero otherwise. It is also necessary to specify in the event of competition which process shall have the use of a crane when it becomes available. This is achieved by associating with each process u a priority number p_u and choose that process from those with non-empty queues and the necessary machines available which has minimum priority number.

A *C*-activity is now necessary. The *B*-activity becomes, using L for a count of cranes:

*B*1

Set name of state and adjust count of cranes.

If a crane was used or no queue for the process add to count of process machines.

Otherwise find an item available in state $j-1$, adjust queue count and commit item to state j. If last state dealt with, print.

Consider next process. If a crane is wanted and none is available, adjust queue count.

Otherwise if no crane required, or if one already idle and a process machine is available, adjust crane count (if required) and machine count and commit to next process.

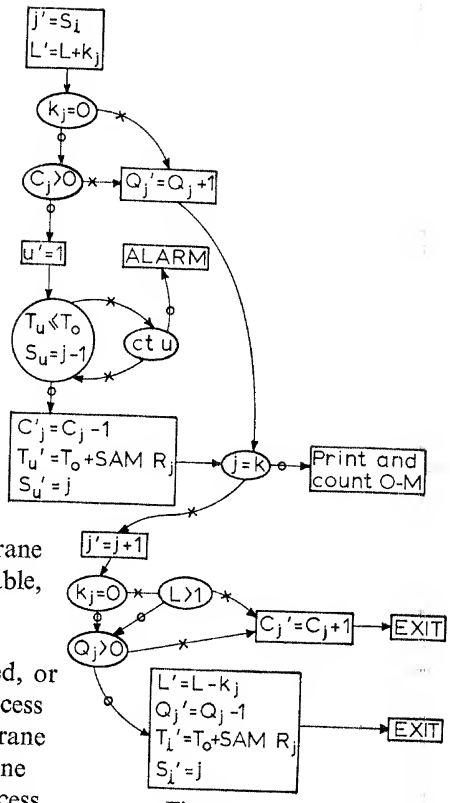


Figure 67

*C*1

If a crane is available, find the process with minimum priority number for which a crane is needed and at least one machine and one item are available.

Find such an item.

Reduce count of cranes, items and machines and commit item to selected process.

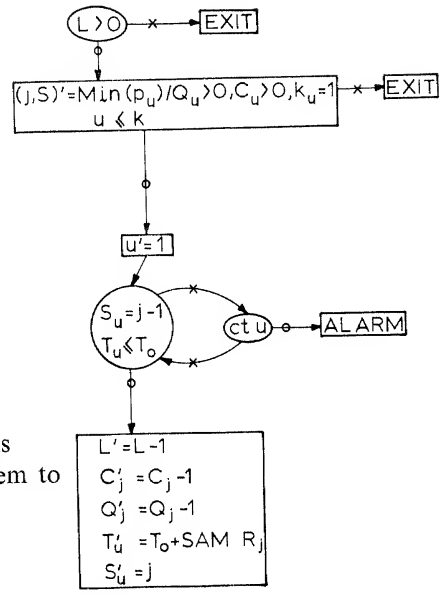


Figure 68

The mechanism for stopping the simulation is now on the count M and, if the duration mechanism is built into some standard structure, the duration must be set very large to prevent it stopping the simulation prematurely. A little care is required with the initial conditions to start the simulation.

If the conventional approach is taken a very long list of machines is produced (of length $\sum n_j$), and as each machine finishes processing an item the long list of items must also be scanned to find a suitable one to process next. If one is not available, a second activity is required to search for a suitable machine when such an item is released by an earlier machine.

The technique used throughout the examples of this and the preceding chapters has been one in which the simulated time is advanced by irregular steps to the next significant event and these steps are found by scanning a relevant set of times. An alternative procedure is to advance time by a regular amount and then determine if any event has occurred in the interval.

This method has the disadvantage that for an equivalent accuracy the number of time-advancing steps is greatly increased without reducing the number of time comparisons made at each advance. Also the possibility of every activity is questioned at each time advance.

The constant time step technique has now been discarded by most workers for the event tracing type of simulation.

It is possible, however, to describe some processes at least approximately by counts of the numbers of events in a fixed time period, and then constant time advance becomes desirable.

A simple queue simulation can be constructed in this manner. Consider a long queue in front of a machine which can deal with X items per hour. The quantity X is considered a random variable, which can be directly observed from the plant. Similarly, the input to the queue can be described by the number of items Y arriving in each hour. Then the throughput in the rth hour is given by the recursion

$$T_r = \min (X_r, Z_r + Y_r)$$

where Z_r is the stock at the beginning of the hour. This simply states that the throughput is the potential throughput or the total available, whichever is the lesser.

The stock at the end of the hour is

$$Z_{r+1} = Z_r + Y_r - T_r = \max (0, Z_r + Y_r - X_r)$$

In these equations X_r, Y_r, are samples from two distributions.

These formulae are only approximate as it is assumed that all the items to arrive will be available when the service needs them. In fact, part of the variability of the number processed arises because delays may be caused to the server by the lack of items to serve. Some allowance is made for this in the distribution of X, but without discrimination. Large values of X should not appear when $Y+Z$ is small, since this is when delays are likely to reduce the throughput.

There are systems where the scheme is quite accurate. Consider a large stock which will never run out, replenished by arrivals and used by two processes in tandem. Assume from the nature of the product that no stock may accumulate between the processes.

In the case when the natural throughput of each process is given by a distribution, then the throughput of the tandem process is

$$\min (X, Y)$$

If the stock is Z and the input is U per hour and the output is T, we have the recursion

$$T_r = \min (X_r, Y_r)$$
$$Z_{r+1} = Z_r + U_r - T_r$$

The difficulty, in this case, is to determine the free distribution of throughput for each process. It requires records showing periods when process 2 was waiting for items (to estimate for process 1) and periods when process 1 was held up continuously by process 2 (to estimate for process 2).

Even in this case, there are end effects between successive hours which are not quite accurately dealt with. An exact number of pieces are rarely dealt with in each interval and the partial process times must be rounded up or down. A little care is needed to prevent bias.

The technique is best used for quick surveys of a situation where the number of items involved in each interval is quite large (say, 20 to 50 items).

CHAPTER 13

DESIGN OF SIMULATION EXPERIMENTS

The preceding chapters have described how to produce efficient computing procedures for simulating a complex industrial plant. Each run of such a procedure has produced a single vector of answers. These can be thought of as a single statistic produced by a complex procedure from the whole set of random variables which represented various time processes involved. Thus the value so produced is liable to fluctuation and repeated runs are necessary to establish the magnitude of this variability and to reduce the sampling error on the overall mean value quoted.

The simple procedure of repeating runs, treating these as independent sample values until variability has been reduced to the desired level, is not usually feasible as the computing labour to obtain a single sample value is quite high and the quantity under examination is often highly variable (in fact, if it were not so, some estimate of its mean value could be found by a deterministic approximation in most cases).

Thus it becomes desirable to utilise the variance reducing techniques that have been considered for Estimation in Chapter 9. The three techniques described there can be listed as the use of:

(i) stratification;
(ii) control variates;
(iii) antithetic variables.

For successful stratification, it is necessary to devise strata for which either the probabilities of a sample of input variables falling into the strata are known, or given that the input variables have so fallen, the conditional expected value of the quantity under estimation can be calculated.

The possibility of determining a partition for which either of these conditions can be met is remote in industrial simulation and will not be further considered. There is, however, one similar device that can occasionally be used. Suppose the simulation involves the use of a set of k similar machines and the rules for using them are symmetrical, i.e. no priorities are involved, and when a choice of machines is available for a job a random selection is made. Suppose moreover that the

utilisation of the machines is low, so that over the duration studied an appreciable proportion of the runs made (from randomised starting conditions) do not use all the machines.

Now suppose that the number of runs involving machines $i_1, i_2, \ldots, i_r$ is $n(i_1, i_2, \ldots, i_r)$ and the average response from these runs is $\bar{Z}(i_1\, i_2, \ldots, i_r)$. Response is used here as a neutral term for the quantity being estimated. It may, for example, be a total delay time on a machine, an output, the maximum length of a queue, the number of delays or the mean delay on a machine.

Now the usual estimate can be written as

$$\sum_p n(i_1, i_2, \ldots, i_r)\bar{Z}(i_1, i_2, \ldots, i_r)/\sum_p n(i_1, i_2, \ldots, i_r)$$

where p is the set of all possible sub-sets of k objects.

However, from the symmetry of the stratification, the probability of falling in the class $(i_1, i_2, \ldots, i_r)$ can be written as p_r depending only on r, the number of machines, and not on the individual names of the machines. This fact can be utilised in the estimation by using the estimate:

$$Z = \left\{\sum_{r=0}^{k} \sum_{pr} n(i_1, \ldots, i_r) \sum_{pr} \bar{Z}(i_1, i_2, \ldots, i_r)\right\} \Big/ \sum_p (n(i_1, i_2, \ldots, i_r))$$

No satisfactory analysis of the reduction of the variance using this estimate seems available and great caution should be used to ensure that the symmetry conditions apply, otherwise bias can be introduced.

The application of the control variate is quite straightforward. Select any time distribution involved in the simulation which clearly affects the response. For example, if the response is the time taken by a simple queue system to deal with a fixed number of items, then an important distribution is the process time distribution. We would expect the average of the process times actually experienced in any run to be positively correlated with the response from that run. As another example, the average cycle time of a furnace will be positively correlated in a simulation run with the time to achieve a given output.

Since the distribution from which these sample variables are drawn is known, the true average value is known and thus the conditions for the control variate techniques are all satisfied.

Moreover, the extra work required to calculate the control variate is quite small and in most cases can be neglected. In this case, the technique always shows an economy.

In many cases, of which the two examples are typical, the natural response is negatively correlated with the control variate. The natural responses in the two examples are output in a fixed time (i.e. proportional to the reciprocal of the response suggested). This can be dealt with by noting that if y and z are negatively correlated $-y$ and z are positively correlated.

Thus the variable estimated is $z+\alpha y$ and α is chosen by the method of least squares. Formally there is no difference and the sign of the estimate of α is an indication of whether y should be regarded as positively or negatively correlated with z.

In experiments of this kind there are often several variates which can be used for control. The control variate technique can be extended. Suppose the response is z and the control variates are $y_1, y_2, \ldots, y_m$. Then the variate estimated is

$$z+\sum_{i=1}^{m} \alpha_i y_i$$

and the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_m$ can be estimated by the method of least squares as before. To avoid bias, the partition of the total sample into k equal parts can be used to prepare k estimates of $\alpha_1, \alpha_2, \ldots, \alpha_m$ independent of the sub-sample means $\bar{z}, \bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k$.

The technique is closely akin to that of the analysis of co-variance used to utilise the known values of concomitant variables. In this analysis, a linear law for $E(z)$ as a function of $y_1, y_2, \ldots, y_k$ is usually assumed, but the analysis of Chapter 8 shows that this is not strictly necessary for the validity of the method. If the linearity does hold the partitioning of the sample is unnecessary. However, in most cases this linearity is most unlikely, and the extra cost of the full analysis is justified.

Turning now to the method of antithetic variates, we note that if an effective control variate has been found, a run in which the samples from the distribution in question have tended to be greater than the true mean (leading to an average greater than the true mean) then the response will tend to be greater than its true mean. If a companion run can be made in which each sample from the distribution is negatively correlated with its mate, then the response from the run will tend to be less than its true mean. Similarly, if the samples from the first distribution tend to be below average. Thus the two responses will be negatively correlated.

It now remains to see how to ensure the negative correlation in the

samples from the control variate. If these are obtained by the inverse cumulative distribution transformation this is easy. If the uniform random variable ξ chosen is greater than a half, the resulting variable is greater than the median (and hence tends to be greater than the mean). On the other hand a uniform variable less than a half leads to a variable less than the median. If the two runs are sampled by using ξ and $1-\xi$ respectively, then the resulting variables are negatively correlated. ξ and $1-\xi$ are negatively correlated with coefficient -1 and the non-linear transformation merely reduces the magnitude of the correlation but it will always remain negative.

The advantage of using pseudo-random numbers for uniform random variables now becomes clear. This not only allows a run to be reproduced by starting each generation at the same initial value but also enables the negative correlation required to be obtained in two 'anti-thetic' runs. Each individual process is provided with its own generator giving uniform random variables $u_1, u_2, \ldots, u_m$ for the m samples used in the run.

In the first run the transformation is effected using $u_1, u_2, \ldots, u_m$ as the independent variables. In the second run the variables used are $1-u_1, 1-u_2, 1-u_3, \ldots, 1-u_m$. All other generations from other distributions are started at the same initial value in the two runs.

In practice, the different times generated may involve a different number of samples being taken on the two runs, but this has no serious effect although it does tend to weaken the correlation effect.

By symmetry we can argue that the variance of the response on these two runs will be equal and use their average as an estimate of the mean response. Even if in some obscure way the assumption of negative correlation is not justified the estimate is still valid.

If a total of $2n$ runs are to be made, suppose n pairs of runs are made yielding estimates z_i, z_i'. The starting conditions for each of a pair are identical, but the different pairs are started with arbitrary conditions including random choice of initial values for each pseudo-random number generator.

The estimate of the response is merely

$$\hat{z} = \frac{1}{2n} \sum_{i=1}^{n} (z_i + z_i')$$

An estimate of variability is required and the best is obtained from

$$s^2 = \frac{1}{2(n-1)} \{\Sigma(z_i - \bar{z})^2 + \Sigma(z_i' - \bar{z}')^2\}$$

This is the appropriate variance for a typical response starting from random conditions. The estimate

$$s'^2 = \frac{1}{4(n-1)} \Sigma\{(z_i+z_i')-(\bar{z}+\bar{z}')\}^2$$

is not valid for this purpose although it is the correct estimate to use when estimating the sampling error in $\hat{z}$.

An extension to utilise several control variates is possible. It is best not to change more than one variable at a time since a second change may weaken the correlation caused by the first one. If there are k distributions $R1, R2, \ldots, Rk$ to be used as sources of antithetic variables we may proceed thus. Choose a sample size $n.2^k$. In each of n sets of 2^k samples allot the sampling procedure as in a 2^k factorial experiment. Associate a vector $v_1, v_2, \ldots, v_k$ with each sample. If $v_i = 0$, treat the distribution R_i normally; if $v_i = 1$ use the variable $1-u_i$ in the transformation to R_i.

For $k = 3$, the following table illustrates the arrangement

SAMPLE		1	2	3	4	5	6	7	8
$R1$	u_1	0	1	0	1	0	1	0	1
$R2$	u_2	0	0	1	1	0	0	1	1
$R3$	u_3	0	0	0	0	1	1	1	1

The pairs (1, 2), (3, 4), (5, 6) and (7, 8) form antithetic pairs from distribution $R1$ and their averages can be denoted by 1′, 2′, 3′, 4′.

The pairs (1′, 2′), (3′, 4′) are antithetic pairs from distribution $R2$ and their averages form an antithetic pair for $R3$. The grand average z is taken as the sample value and the final estimate is taken as the grand average of the n such values. The estimated sampling error is found from the mean square deviation of the z's but the expected variability of the response on the plant itself is derived from the mean of the 8 mean square deviations of the sets of n commonly treated runs.

Since n should be kept fairly high so that this latter estimate is not too imprecise, the size of k is restricted to 2 or 3 in practice.

The control variate technique may be applied to the combined variable z using either the antithetic distribution or preferably some other.

If a fixed accuracy of estimate is required, the sequential procedure given in Chapter 8 can be used.

Even if in any particular case, these techniques do not cause any

drastic reduction in variation, they cost so little to implement that they are always to be recommended.

The methods of generating starting conditions merit some comment. In a complex simulation it is very difficult to concoct sensible starting conditions if these are intended to be typical of the state of the plant when inspected at random. If the plant closes down at frequent intervals (each day or each weekend) then the real starting conditions are easy to determine and each run can be started from there. But if the plant runs are very long a start of the simulation from 'typical' conditions is required.

There can be no general theory of how to do this but the technique most favoured is to invent starting conditions and allow the simulation to proceed for some time and take the final conditions as the initial conditions of the genuine run. This raises the question of how long to make the preliminary run and again no definite answer is possible. A general requirement is that the longest cycle in the plant should have been executed at least 3 or 4 times before transient abnormal behaviour induced by non-sensible starting conditions can be expected to have died away.

It also needs a decision whether successive runs shall consist of wholly independent runs (started as described above) or shall be obtained by using the final calculation of one run as the start of the next. This should certainly not be done if real plant runs are only short and the continued running of the plant may reveal instabilities that are never given time to develop in the real plant. For example, a queue may operate with input faster than average output but this will not matter if the remaining items at plant close down are dealt with differently during the shut-down.

The theoretical solution to this problem depends on whether the response in one period is positively or negatively correlated with that in the next period. In the former case, separate starts should be used in the interest of variance reduction. In the latter case, continued runs are better. In practice, the labour of making valid fresh starts weighs heavily in favour of continued runs.

An allied question is the length of time that should constitute a run. This should be related to the purpose of the simulation and is rarely settled by statistical considerations. A plant with short real runs should be simulated for complete runs so that end effects are consistently dealt with. Plant with long runs can be dealt with by shifts, days or weeks. In case of doubt as to which is an appropriate duration, the

shorter should always be chosen as results can always be amalgamated but not broken down. The extra printing involved should not be an intolerable burden. If it appears so, the responses should be stored in the computer and re-processed to give a small amount of output.

So far it is assumed that the purpose of the simulation is to estimate some quantity as accurately as possible. In the first stages of an investigation this is necessary to test a simulation of an actual plant as a means of verification of the validity of the model. However, a model that exactly reproduces an existing plant is of little value (except possibly for educational and training purposes), since the plant itself can give all the information that the model can.

The second stage of the investigation is to change the model to represent a proposed change of the real plant. The purpose of the simulation runs then is to measure the effect of this on the response. The interest now shifts from absolute values to comparisons and we accordingly turn to the problem of designing the runs to enhance the accuracy of such comparisons.

Consider the simple situation in which two models are to be compared. It is assumed that the models differ only in some small part (even though we hope it to be an important difference!) and that most of the input distributions are the same. The usual differences are either in the number of ancillary machines offering service or in the rules of behaviour of the machines under different circumstances.

Clearly two runs, one on each model, are likely to give a better estimate of the difference between the models if initial conditions and process times are the same than if not.

This pair of runs is made from the same initial conditions and using the same initial pseudo-random numbers for the distributions (if one model contains extra or different distributions a random choice is made for these). The difference of the responses is taken and this is treated as a response and the pair of runs treated as a single run. Both the control variate technique and the antithetic treatment can be used and sequential estimation used, if desired.

As before, if variance reducing techniques are used, careful distinction must be made between the sampling error of the estimate and the variance of the statistic in the population.

The same process can be applied if three or more different models are to be tested.

If there are t models then n sets of t runs are made, each member of each set starting from common initial conditions and using a different

one of the k models. Differences between averages for the n runs on each model give estimates of the model differences in which a large amount of the variability has been removed.

The number of sets, n, should be kept as large as possible in order that the differences are measured over as widely a differing set of conditions as possible and to enable the population variability of the responses to be measured. If t becomes large, this implies a very large number, nt, runs and this may be prohibitive of labour.

The solution then lies in using n incomplete sets of runs—incomplete in the sense that not all k models are represented in each set. In fact, we adapt the theory of block experiments to the situation. The blocks consist of the sets of k runs with common initial conditions, the treatments are the t different models and the plots are the individual runs.

Our treatment of the case $t = k = 2$ is a paired comparison experiment. The case of small $k = t$ is dealt with by a complete randomised block experiment. For large t we must use incomplete block experiments.

Randomisation would play no useful role in the experiments as so far described since each run of a set will produce an identical response if 'treated' with the same model. The purpose of randomisation is to make that true of the expected response. There may be some merit in not making the initial conditions of the runs in a set identically equal but allowing some minor variations so that the assumptions of the analysis technique used for the experiments are approximately valid.

The whole repertoire of designs available for real experiments is now open to us in our simulation experiments. The whole subject is too large to be dealt with here and the reader must look elsewhere for more detailed information.

However, one design is so useful that a brief description is desirable. This is a special case of a balanced incomplete block design. In general this is a design arranged so that every model is run equally often and every pair of models are run together in the same number of blocks. The simplest case of this is when $k = 2$ and each block is just one pair. Since there are ${}^kC_2$ pairs the number of blocks is a multiple of this. For example with 5 models (labelled 1, 2, 3, 4, 5) the 10 blocks of a single replicate consist of the pairs (1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5). At least two replicates are desirable if the population variability of any difference is to be estimated. The analysis will ordinarily assume equal variability of response in all

models and hence in all differences. To test this would require many replicates.

This design and the extension of it, known as partially balanced designs, assume no structure to the set of models under test. However, often this set has been generated in some systematic way. For example, the models may be developed from the real plant by adding a storage tank of capacities 10, 15, 20, 25 units, by adding 1, 2, 3 extra cranes in a machine shop, or by trying two variants of a queue discipline on each of three queues (8 models in all).

In the first examples, the ultimate purpose will be to draw a graph of response against an independent variable. In the last example, we have a factorial arrangement of 3 factors (queue) at 2 levels. This is a very common structure to experiments and the theory of the design of experiments has concentrated a good deal of attention on the problem of designing new experiments.

Once again the reader must look elsewhere for details.

In any experiment with t models, although there are $\frac{1}{2}t(t-1)$ differences that can be studied, only $t-1$ of these are independent and there is some skill in choosing the appropriate combinations. For example, if z_1, z_2, z_3 are the three responses with 1, 2, 3 extra cranes, the component $\frac{1}{2}(z_3-z_1)$ measures the average increase in response per extra crane while $z_3-2z_2+z_1 = (z_3-z_2)-(z_2-z_1)$ measures the change in increase of response as the number of cranes increase.

The independence of the $t-1$ differences is in an algebraic sense. This is quite distinct from the statistical dependence or correlation between *estimates* of various differences. It is the distinguishing feature of complete block designs that algebraic independence implies statistical independence (at least under some mild assumptions). All incomplete block designs give correlated estimates of responses. Of course, it is possible to determine special comparisons (of differences) that are independent, but unless the experiment has been specially designed to achieve this, they are unlikely to be comparisons that are of interest.

The whole field of experimental design for simulation experiments is in its infancy and offers a fertile field for further research.

REFERENCES

F. J. Anscombe (1953), Sequential Estimation, *J. Roy. Statist. Soc.* A, XV, No. 1.

S. Beer (1957), The Mechanical Simulation of Stochastic Flow, *Proc. First Inter-Conf. ORSA*, Baltimore, 166–75.

BISRA Publication (1959), *Symposium on Computer Simulations*, Sheffield.

E. Bofinger and V. I. Bofinger (1958), A periodic property of pseudo-random sequences, *J. Assoc. Comp. Mach.*, 5, 261–5.

G. E. P. Box and M. E. Muller (1958), A note on the generation of normal deviates, *Ann. Math. Stat.*, 28, 610–11.

S. W. Broadhurst and D. T. Harmston (1950), An Electronic Traffic Analyser, *Post Office Electrical Engineer's Journal*, 42, 181.

J. C. Butcher (1961), Random Sampling from the Normal Distribution, *The Computer Journal*, 3, No. 4.

J. W. Butler (1956), Machine Sampling from given Probability Distributions, *Symposium on Monte Carlo Methods*, 249, New York: Wiley.

J. Certaine (1958), On Sequences of Pseudo-Random Numbers of Maximal Length, *J. Assoc. Comp. Mach.*, 5, 353.

J. M. Cook (1957), Rational formulae for the production of a spherically symmetric probability distribution, *Math. Tables Other Aids Comp.*, 11, 81.

R. R. Coveyou (1959), Serial Correlation in the Generation of Pseudo-random numbers, *J. Assoc. Comp. Mach.*, 6, 72–4.

D. W. Davies (1954), Discussion on Symposium on Monte Carlo Methods, *J. Roy. statist. Soc.* B, 16, 1.

P. Davis and P. Rabinowitz (1956), Some Monte Carlo experiments in computing multiple integrals, *Math. Tables Other Aids Comp.*, 10, 1–7.

H. J. A. Duparc, C. G. Lekkerkerker and W. Peremans (1953), Additive Congruential Methods for Pseudo-Random Numbers, *Math. Centrum.*, Amsterdam.

A. R. Edmonds (1959–60), The Generation of Pseudo-Random Numbers on Electronic Digital Computers, *Computer Journal*, 2, 181–5.

W. Feller (1950), *An Introduction to Probability Theory and its Applications*, Vol. 1, 1st ed., New York: Wiley.

E. C. Fieller and H. O. Hartley (1954), Sampling with Control Variables, *Biometrika*, 41, 494–501.

R. A. Fisher and F. Yates (1948), Statistical Tables, Oliver & Boyd, 104–9.

G. E. Forsythe (1951), Generation and testing of random digits at the National Bureau of Standards, Los Angeles, *Monte Carlo Method*, National Bureau of Standards AMS 12 (U.S. Government Printing Office, Washington, D.C.), 34–5.

I. J. Good (1953), The Serial Test for Sampling Numbers and other tests for randomness, *Proc. Camb. Phil. Soc.*, 49, 2, 276–84.

R. E. Greenwood (1955), Coupon collectors' test for random digits, *Math. Tables Other Aids Comp.*, 9, 1.

M. Greenberger (1961), An *A Priori* Determination of Serial Correlation in Computer Generated Random Numbers, *Math. Comp.*, 15, 383–9.

F. Gruenberger and A. M. Mark (1951), The d^2 Test of Random Digits, *Math. Tables Other Aids Comp.*, 5.

J. H. Halton and D. C. Handscomb (1957), A Method for Increasing the Efficiency of Monte Carlo Integration, *J. Assoc. Comp. Mach.*, 4, No. 3.

P. C. Hammer (1951), The mid-square method of generating digits, *Monte Carlo Method*, National Bureau of Standards, AMS 12 (U.S. Government Printing Office, Washington, D.C.)

J. M. Hammersley and J. G. Mauldon (1956), General Principles of Antithetic Variates, *Proc. Camb. Phil. Soc.*, 52, 3, 476–81.

J. M. Hammersley and K. W. Morton (1956), A New Monte Carlo Technique: Antithetic Variates, *Proc. Camb. Phil. Soc.*, 52, 3, 449–75.

J. M. Hammersley and K. W. Morton (1954), Poor Man's Monte Carlo, Symposium on Monte Carlo Methods, *J. Roy. Statist. Soc.* B, 16, 23–38.

C. Hastings (1955), Approximations for Digital Computations, Princeton University Press, Princeton, N.J.

R. K. Hayward and E. L. Bubb (1957), E.R.N.I.E., *Post Office Electrical Engrs' J.*, 50, 1–6.

R. K. Hayward, E. L. Bubb and H. W. Fenson (1957), Computer Selects Premium Bond Winners, *Electronics*, 30, 7, 138–43.

J. S. Hicks and R. F. Wheeling (1959), An Efficient Method for Generating Uniformly Distributed Points on the Surface of an *n*-dimensional Sphere, *J. Assoc. Comp. Mach.*, 2 (4).

D. L. Johnson (1956), Generating and testing Pseudo-Random Numbers on the IBM Type 701. *Math. Tables Other Aids Comp.*, 10, 8–13.

M. L. Juncosa (1953), Random number generation on the BRL high-speed computing machines, Report No. 855, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland.

H. Kahn (1949), Modification of the Monte Carlo Method, *Scientific Computation Seminar Proceedings*, IBM Applied Science Dept.

H. Kahn (1954), Revised 1956, Applications of Monte Carlo, *RAND Project Report*, RM-1237-AEC.

M. G. Kendall and B. Babington Smith (1938), Randomness and random sampling numbers, *J. Roy. Statist. Soc.*, 101, 147–66.

M. G. Kendall and B. Babington Smith (1939), Second paper on random sampling numbers, *J. Roy. Statist. Soc.* Suppl., 6. 51–61.

W. O. Kermack and A. G. McKendrick (1937), Tests for Randomness in a Series of Numerical Observations, *Proc. Roy. Soc. of Edinburgh.*

W. O. Kermack and A. G. McKendrick (1937), Some Distributions associated with a randomly arranged set of numbers, *Proc. Roy. Soc. of Edinburgh.*

D. H. Lehmer (1954), Description of 'Random Number Generation on the BRL high-speed computing machines', *Math. Rev.*, 15, 559.

D. H. Lehmer (1951), Mathematical methods in large-scale computing units, *Proc. Second Symposium on Large-scale Digital Calculating Machinery*, Harvard University Press, Cambridge, Mass., 141–6.

J. R. Lion and K. S. Lion (1959), Electric System for Random Selection of Two Alternate Circuits, *Rev. of Scientific Instruments*, 30, 4.

P. C. Mahalonobis (1934), Tables of random samples from a normal population, *Sankhyā*, 1, 289.

G. Marsaglia (1961), Expressing a random variable in terms of uniform random variables, *Ann. Math. Stat.*, 32, 3, 894–98.

G. Marsaglia (1961), Generating exponential random variables, *Ann. Math. Stat.*, 32, 3, 899–900.

N. Metropolis and S. Ulam (1949), The Monte Carlo Method, *J. Am. Statist. Assoc.*, 44, 335–41.

J. Mosham (1954), The generation of pseudo-random numbers on a decimal calculator, *J. Assoc. Comput. Mach.*, 1, 88–91.

M. E. Muller (1958), A Comparison of methods for generating normal deviates, Technical Report No. 9, Statistical Techniques Research Group, Department of Mathematics, Princeton University.

M. E. Muller (1959), A note on a Method for Generating Points Uniformly on *N*-Dimensional Spheres, *J. Assoc. Comp. Mach.*, 2 (4).

M. E. Muller (1958), An inverse method for the generation of random normal deviates on large-scale computers, *Math. Tables Other Aids Comp.*, 12, 167–74.

NBS, *Applied Mathematics Series 23*, Tables of Normal Probability Functions, U.S. Government Printing Office, Washington, D.C. (1953).

J. von Neumann (1951), Various Techniques Used in Connection with Random Digits, '*Monte Carlo Method*', *Nat. Bur. Standards Applied Math. Series 12*, 36–8.

E. S. Page (1954), Monte Carlo Methods for the solution of some integral equations, *Proc. Camb. Phil. Soc.*, 50, 414.

A. Rotenberg (1960), A New Pseudo-Random Number Generator, *J. Assoc. Comp. Mach.*, 7, 1, 75–7.

F. Sterzer (1959), Random Number Generator using Subharmonic Oscillators, *Rev. of Scientific Instruments*, 30, 4.

Student (1908), On the probable error of a mean, *Biometrika*, 6, 1.

O. Taussky and J. Todd (1956), Generation and testing of pseudo-random numbers, *Symposium on Monte Carlo Methods*, University of Florida, Wiley, 15–28.

D. Teichroew (1953), Distribution sampling with high speed computers, Ph.D. Thesis, University of North Carolina.

The Rand Corporation: *One Million Random Digits and 100,000 Normal Deviates*, The Free Press, Glencoe, Illinois.

W. E. Thompson (1958), A modified congruence method of generating pseudo-random numbers, *The Computer Journal*, 1, 83.

W. E. Thompson (1959), ERNIE—A Mathematical and Statistical Analysis, *J. Roy. statist. Soc.* A, 122, 3.

L. H. C. Tippett (1927), Random Sampling Numbers, *Tracts for Computers*, XV, Cambridge University Press.

L. H. C. Tippett (1925), *Biometrika*, 17, 364–87.

K. D. Tocher (1954), The application of automatic computers to sampling experiments, *J. Roy. statist. Soc.* B, 16, 39.

J. W. Tukey (1934), On the distribution of the fractional part of a statistical variable, *Mathematical Abstracts*, Academy of Sciences, U.S.S.R., Moscow.

S. Ulam (1951), On the Monte Carlo Method, *Proc. of a Second Symposium on Large-scale Digital Calculating Machinery*, 207.

V. W. Vickery (1938–9), On Drawing a Random Sample from a set of Punched Cards, *J. Roy. statist. Soc.* Supplement, 5–6, 62–6.

D. F. Votaw, Jnr., and J. A. Rafferty (1951), High Speed Sampling, *Math. Tables Other Aids Comp.*, 5, 1–8.

G. M. White (1959), Electronic Probability Generator, *Review of Scientific Instruments*, 30, No. 9.

H. Wold (1954), Random Normal Deviates, *Tracts for Computers*, XXV, Cambridge University Press.

INDEX